PRECURSOR

HERE IS THE STRONGEST LABOR WARNING EVER.

The process of making a Vocaloid DB is not only tiresome, jarring and irritating. It's generally convoluted, and extremely finicky. There is no current shortcuts, or exclusions. You either do it, or you don't.

If you don't have the time, the patience, or the space on your computer. (5+GB) Then you're better off just not even attempting.

This is going to take a lot of effort.

You cannot skip, or skim over a single detail when reading.

(If you do and join the discord server without reading, we will not help you.)

If I did not mention it in this guide, do not touch it.

VOCALOID SVS TUTORIAL

A step-by-step guide on creating a Voicebank for Vocaloid3-5.

Written by: EnlilGen

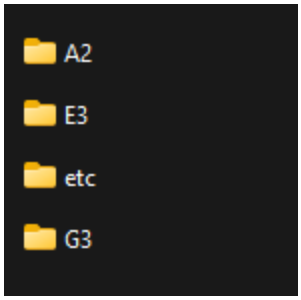
DATASET PREPARATION

Step 0 - Folder Prep(Optional)

Folder prep is totally up to you. Porting an UTAU CV, CVVC, or VCV is not impossible at all. It's just a matter of what samples you do and don't have. You can just fake a lot of them. For example, a lot of Miku and co's VC are fake. Most V3+ have half hazardous VCV, so I'm sure you'll be fine. Again, however… Refer to the labor warning.

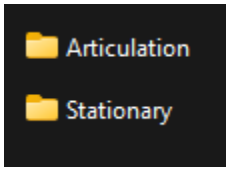
Data for a Vocaloid is no different from regular UTAU data. To start, you would section off your pitches, creating a suitable range for your bank. There are some things to take into account, however. When a pitch is too high, or too low. **Vocaloid tends to give the higher pitches the entirety of the bank.** Pushing your supporting pitches to the background. **So it's either your Suffix (Tone tag) it, or make an entirely new bank.** It's best to do this for things like falsetto, or extreme head voice. One off thing can destroy your voicebank entirely.

Folder Example:



It is inside these folders where you'd make other folders. **Such as S (Stationary), A (Articulation), T (Triphoneme)**, and etc (For one off things).

Folder Example:



Step 1 - Samples

Now you have these pitches, and these folders. (Don't worry, the tool acknowledges subfolders.) What do these even mean and why are they important? Below are definitions.

The V Samples or, Stationaries. The looped sample: This is a short, but long enough sound for the middle of vowels (or consonants) that is looped in the engine. Below is a chart of connections and what they look like.

_a	[a]
_i_m_u	[i] [m] [M]
_e_o_ng	[e] [o] [N]
_l_n_l	[i] [N\] [l]

(This is not like UTAU's resample; the default resampler stretches the middle of notes, not loops.) Without Stationaries, Vocaloids will fade out in the middle of singing a note! You should record a long enough sample to where the resampler doesn't repeat too harshly. You can take them from any sound, really. For the best results, make sure you use the same sample for the opening and close of your vowels. (Note that; whatever you give a stationary to will gain some kind of pitch. Like for example, adding one to 4 phonetic, will give a more melodic sound.)

To pull them up you would go to Viewdb>Stationary

***ADVICE:** take a piece of the sample you find to be stable. Cut it out and stretch it in melodyne.

The C-V V-C Samples or Articulations. The connection samples: Like any concat synth, You need connections of every kind. Below is a chart of connections and what they look like.

_ka_ka_n_ka	[Sil k a k a N k a]	[Sil k] [k a] [a k] [N k]
_ba_bi_bu	[b a b' i b M]	[b a] [a b'] [b' i] [i b] [b M]
_star_street	[Sil s t Q@ s t r i: t]	[Sil s] [t Q@] [Q@ s] [s t] [t r] [r i:]
_kwang_kung	[k w a N k M N]	[k w] [w a] [a N] [N k] [k M] [M N]

For your select language you'd need one of every connection. (C-C connections aren't needed but helpful for making transitions smoother.) **Once you have an opening (Sil X) of anything, plus a connection. That is enough for a plug to work.** You don't need endings, but they help. C-V V-C, while a lot easier and fairly light on workflow, is very tedious. For a test bank it's a great idea, but for overall great sound is lackluster.

To pull them up you would go to Viewdb>Articulation
***ADVICE:** Use just enough of the sample, and rely off stationeries for tone.

The V-C C-V samples or Triphonemes. The transition smoother: Again like any concat synth, This is more a cosmetic design and puts a ton of work on your plate. Below are examples of connections and what they look like.

_ka_ka_n_ka	[Sil k a k a N k a]	[Sil k a] [a k a] [N k a]
_ba_bi_bu	[Sil b a b' i b M]	[Sil b a] [a b' i] [i b M]
_star_street	[Sil s t Q@ s t r i: t]	[s t Q@] [s t r]

You can really do this like a regular voicebank, and since it's purely cosmetic. You can choose where and how to add them. You wouldn't do them for V-V connections, but it is entirely possible. It will help smooth out whatever you decide is an undesirable connection while also adding a ton of variation. If you're not up to it, don't do it. It becomes very taxing after a bit because of all the window switching. It also doesn't have to be one or the other. You can do Articulations and Triphonemes on the same file.

To pull them up you would go to Viewdb>Triphone
***ADVICE:** Doing [Sil C V] or [C C V] samples would greatly improve your voicebanks quality.

There's such a thing as a **nonsense connection**. Below are examples.

_dream	[dh r i: m]	[r i: m]
_ka_ka_k	[k a k a k Sil]	[k a k] [a k Sil]
_star_street	[Sil s t Q@ s t r i: t]	[t Q@ s]
_star_street	[Sil s th Q@]	[s th Q@]

You have no use for ending C-C. The C-V V-C can fill that gap, and then making connections with C's at the end of a triphoneme won't really help it with sound. You're adding pointless data, a nonsense connection. Also creating triphone connections with the wrong phonetics, such as th (Used for aspiration) defeats the point of the phonetic in general.

Vocaloid phonemes also take into account context no matter the language. So when recording be sure to take those phonetics into account. Here's some examples.

Senpai	s e m p a i
Kinko	k' i N k o
Konya	k o N\ j a
Sanban	s a m b a n
Zanbi	dz/z a m' b' i
Konnichiwa	k o J J i tS i w a
Lovely	l0 V v l i:

Sometimes these samples also have stationeries, they're more than likely a nasal, or liquid. Always check other voicebanks for confirmation.

The DBTool reads.**WAVs at 44100 Htz 16bit MONO.** You can do whatever to your audio, it's just like UTAU. **This next step is optional, but for better sound quality;** Throw all of your samples into audacity. Do the noise removal for background noise, Filter Curve EQ> Preset> Low rolloff for Speech, (If you have a tool for de-essing that's even better.) Then normalize the audio.

Don't save over the original samples, rather make a new folder with your samples that have been filtered. Export them all using batch export, doing by the track names which should be the file names. (It's also completely okay to pitch correct your samples.)

Step 2 - Transcription files

It's just what it sounds like. The transcript file is how the DBTool recognizes phonetics in the audio. You'd type out what's in the audio file in the preferred dictionaries phonetic system, while also making **[brackets]** underneath. Brackets are the instructions for the fragments of audio you want pulled from the samples. (Think of them as UTAU Aliasing)

Stationary transfiles can only have one thing in their brackets **([V]/[a])**, but you can still do multiple brackets. Meaning, you can pull multiple stationeries from one sample.

_a_i_u_e_o_n	Sil a i M e o N\	[a] [i] [M] [e] [o] [N\]

Let's make something clear about [brackets]. **Brackets only care about what's first.**

_b_ey_b_ey_b	[Sil b eI b eI b Sil]	[Sil b eI] [eI b] [b eI] [eI b]

The issue here is that: **The first [b eI] and [b eI] are the same thing because they both occur first. Leaving the second b eI to be useless!** That's how the program sees it at least. To fix this you'd do

_b_ey_b_ey_b	[Sil bh eI b eI bh Sil]	[Sil bh] [bh eI] [eI b] [b eI] [eI bh] [bh Sil]

Just for the distinction of what is what. This is exactly the same for triphones. For examples

_ba_ba_bu_ba	[Sil b a b a b M b a]	[Sil b a] [a b a] [a b M] [M b a]
_tri_tru_tra	[th r i: th r u: t r V]	[th r i:] [th r u:] [t r V]

With the opening V/C is enough of a distinction.

To make transfiles you can just rename a .txt to .trans, making sure they share the audio file's name. You can also use Canned-Bread's Swiss army knife to quickly make them. They'll name themselves after your samples and you can get to work. If you intend to do only CVVC it'd be okay to use the auto_transcript feature. However, from my personal experience it isn't perfect. Do be sure to clean your transcript, making sure your phonetics are completely correct.

Once generated they are editable in the DBtool, or in a text editor. It's better to take care of them first in a text editor. Batch editing, so struggling later doesn't happen. This speeds up the workflow by a mile. (Notepad++ is a lifesaver as well <3. You can open transfiles in a regular notepad or notepad++ doesn't really matter. They're just text files.)

Some rules of thumb to make this easier

1. **Sil must always have a capital S!** Or else it won't be read by the program at all. Causing you to have to completely redo your segmentation.

2. You don't need to completely transcript the whole sample, or add the Sil's all the way through. **You can record n ka, only transcript N k, bracket [N k], add it, then go on your way.** For any sample or segment, you can just do the part you need and dip. Sure you waste a lot, but this gets things done.

3. **This is for japanese only.** A GRAPH FOR N: **(Yes you need all of them, yes these have stationaries)** For other contextual things, be sure to check your language of choice.

[m]	[b] [p] [m]
[m']	[b'] [p'] [m']
[N\]	[a] [i] [M] [e] [o] [n] [N] [N'] [J] [m] [m'] [h] [C] [h\] [p\] [p\'] [w] [v] [v'] [j]
[N]	[k] [g]
[N']	[k'] [g']
[J]	[J]
[n]	[n] [t] [t'] [ts] [tS] [d] [d'] [dZ] [dz] [4] [4'] [Z] [z] [m] [m'] [J] [N] [N'] [N/]

4. The only consonants that will need Stationaries are the nasals, for English liquids (L) will also need it. Some nasals and liquids should have them no matter the language. **Consonants will not transition smoothly with other consonants if there's no CC Articutions. Basically no connection, no transition.**

5. It's best not to go too hard on pitches. As it can break the Vocaloid. **In general limit your pitches to 3-10.** XSY is completely usable so do not fear! If you want to, you can 3 pitches of articulations, and 5 pitches or more of stationeries. ***You cannot cross two different language/format banks. CVVC only banks don't work with XSY. (To circumvent this just do one VCV sample during development, Doesn't matter the language.)**

6. **Sil C markers should have no sound in them, while Sil C V should have sound in their markers.** It can create a weird drawn out effect if not handled correctly. Liquids, Nasal, and Breaths fall under this too. **The program will not accept [Sil Sil].**

7. **Remember, the languages available in Vocaloid are the only ones with working word dictionaries.** So it's best to work within their limits or, work with phonetics. (You can make a custom one, but it'll be really rough.)

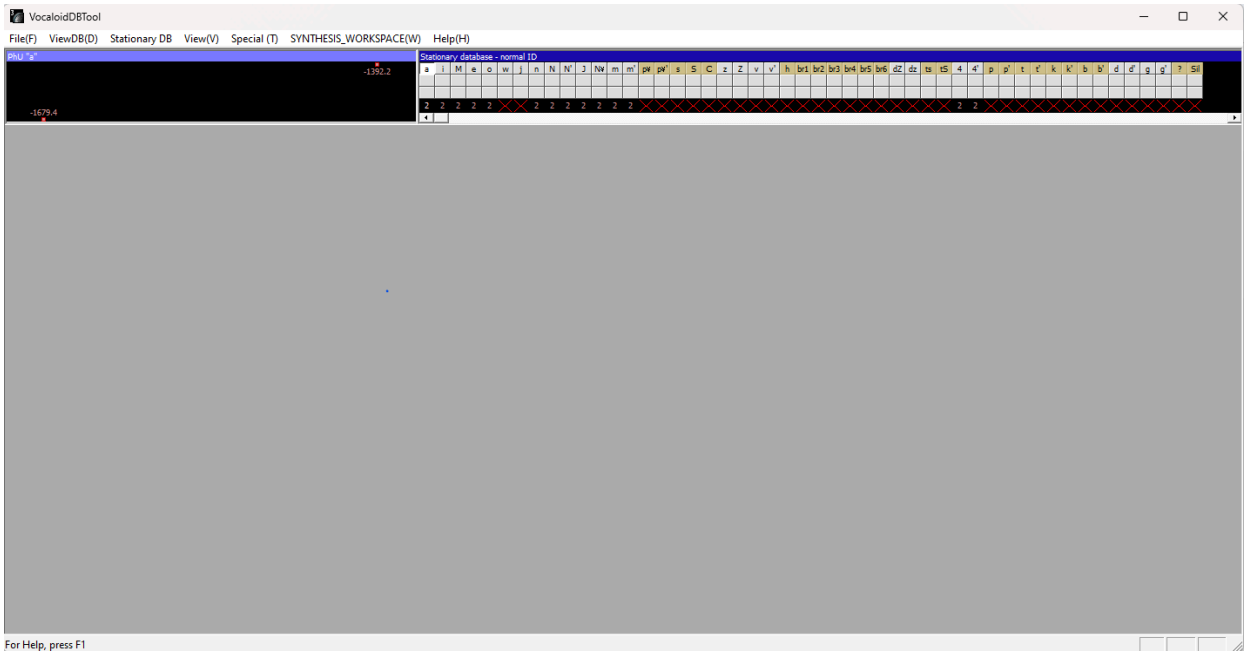
8. Just so you know a bi bi may read as [b' i] [i b'] [b' i] **but you still need a [i b] and [b i] articulation**, That's for all phonetics. For example an [i h] is needed to transition [i ha]. Even if it's [i C] or [C i] the [i h] and [h i] should work no differently.

9. It doesn't hurt to just **copy and paste your samples.**

Step 3 - New Singer Database

After you have everything ready. You can go ahead and make a blank txt file and name it after your singer. (Or codename it doesn't really matter.) You're going to also need a dictionary for the DBTool. They're are some in the folder with the DBTool. It's completely possible to edit them while they're text files. You can also use the dictionary feature in the DBTool to just add things in later. (Explained below)

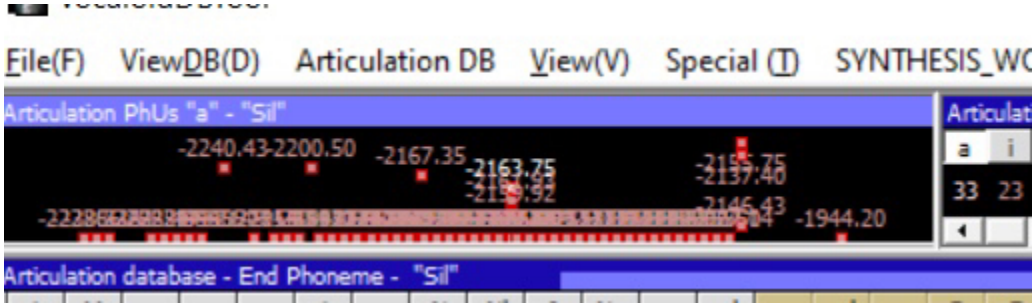
Open up the DBTool, **select File -> New Singer Database.** From there you load your dictionary of choice, and then your blank text document named after your singer. You can click on **ViewDB -> stationary** to see if it worked. You should see all of your phonetics and **X**'s under them. Because there are no samples.


VocaloidDBTool
File(F) ViewDB(D) Stationary DB View(V) Special (T) SYNTHESIS_WORKSPACE(W) Help(H)
Stationary database - normal ID
For Help, press F1

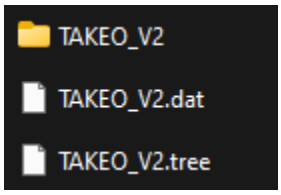
The **X** will change to numbers after samples are added. That goes for all sections. **A red dot** with the frequency of the sample will appear. **From left to right is lower pitch to higher pitch.** You can click on a red dot to view the sample and edit the markers. The same way you do in segmentation. Just make sure to register it by going

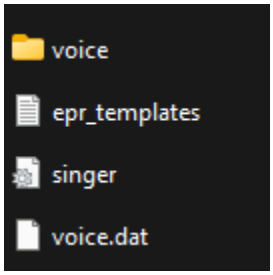
to nameDB > update markers. (If you're getting a weird noise your marker is probably placed wrong.) **When checking your samples make sure you click on the red dot everytime you change samples, Or it wont bring up the correct one.**

Overlap is detrimental to a voicebank. If you add two samples of the same pitch and same phonetic to a voicebank. Vocaloid will switch between the two in a strange fashion. It will sound clunky and uneven, almost sometimes completely destroying your Vocaloid. Please be careful where you are pulling your samples from and what is being taken from where!



After your singer is created, a folder will auto-gen. This will create a folder named after the txt file you called your singer. **A (name).dat, and (name).tree** will appear outside the folder. They are important, do not delete them. When you close the DBtool you can load the bank again from Inside of the folder with the file named **singer.inf.** Simply open the DBtool, load the singer and you're fine.



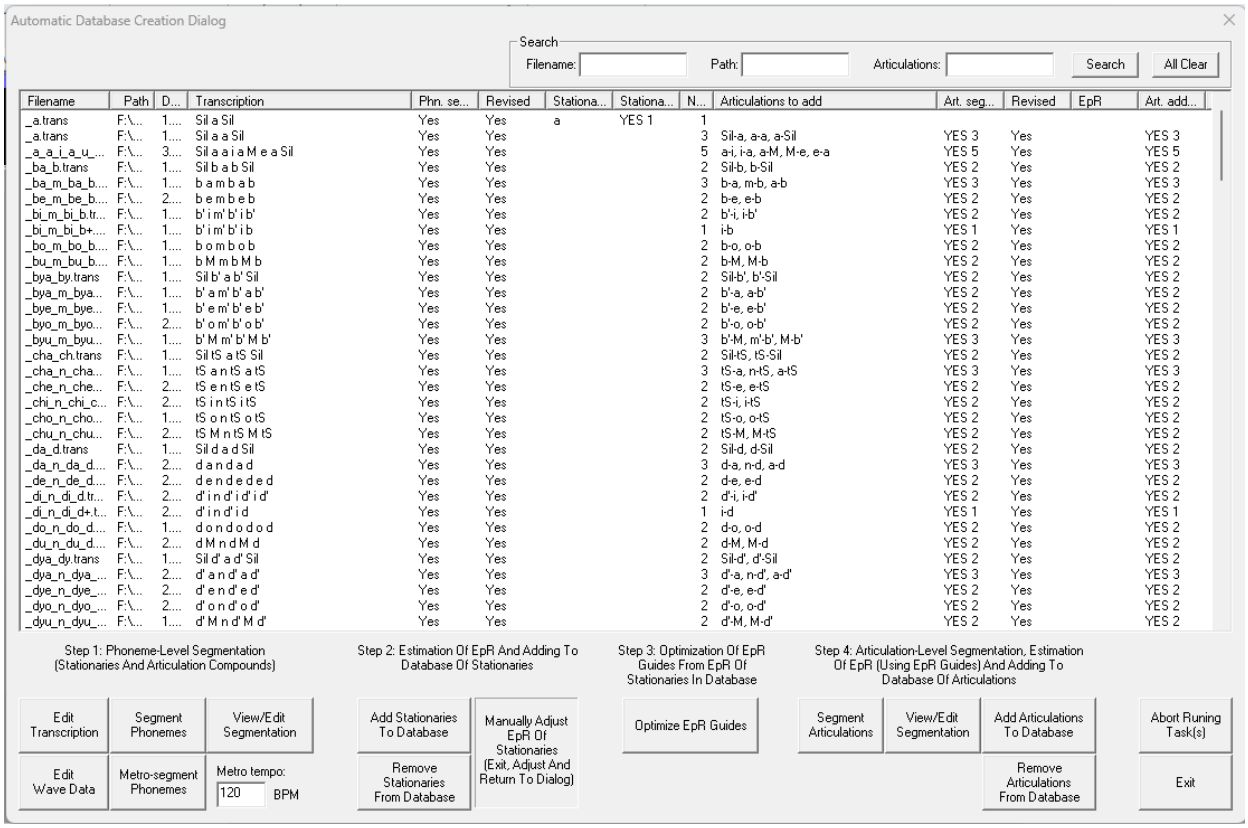


Let's move on to labeling.

VOICE CREATION

Step 0 - Automatic Segmentation

To begin segmentation you would go **File->Automatic Segmentation.** Locate your audio/trans files. Both have to be in the same folder. The DBtool acknowledges subfolders, So you'd be completely fine loading the whole bank. Then locate the DBTool as your segmentation tool. If you did everything right, you should see this.


Automatic Database Creation Dialog
Search
Filename:
Path:
Articulations:
Search
All Clear
Filename | Path | D... | Transcription | Phn. se... | Revised | Stationa... | Stationa... | N... | Articulations to add | Art. seg... | Revised | EpR | Art. add...
_a.trans F:\... 1.... Sil a Sil Yes Yes a YES 1 1
_a.trans F:\... 1.... Sil a a Sil Yes Yes 3 Sil-a, a-a, a-Sil YES 3 Yes YES 3
_a_a_i_a_u_... F:\... 3.... Sil a a i a M e a Sil Yes Yes 5 a-i, i-a, a-M, M-e, e-a YES 5 Yes YES 5
_ba_b.trans F:\... 1.... Sil b a b Sil Yes Yes 2 Sil-b, b-Sil YES 2 Yes YES 2
_ba_m_ba_b.... F:\... 1.... b a m b a b Yes Yes 3 b-a, m-b, a-b YES 3 Yes YES 3
_be_m_be_b.... F:\... 2.... b e m b e b Yes Yes 2 b-e, e-b YES 2 Yes YES 2
_bi_m_bi_b.tr... F:\... 1.... b' i m' b' i b' Yes Yes 2 b'-i, i-b' YES 2 Yes YES 2
_bi_m_bi_b+.... F:\... 1.... b' i m' b' i b Yes Yes 1 i-b YES 1 Yes YES 1
_bo_m_bo_b.... F:\... 1.... b o m b o b Yes Yes 2 b-o, o-b YES 2 Yes YES 2
_bu_m_bu_b.... F:\... 1.... b M m b M b Yes Yes 2 b-M, M-b YES 2 Yes YES 2
_bya_by.trans F:\... 1.... Sil b' a b' Sil Yes Yes 2 Sil-b', b'-Sil YES 2 Yes YES 2
_bya_m_bya... F:\... 1.... b' a m' b' a b' Yes Yes 2 b'-a, a-b' YES 2 Yes YES 2
_bye_m_bye... F:\... 1.... b' e m' b' e b' Yes Yes 2 b'-e, e-b' YES 2 Yes YES 2
_byo_m_byo... F:\... 2.... b' o m' b' o b' Yes Yes 2 b'-o, o-b' YES 2 Yes YES 2
_byu_m_byu... F:\... 1.... b' M m' b' M b' Yes Yes 3 b'-M, m'-b', M-b' YES 3 Yes YES 3
_cha_ch.trans F:\... 1.... Sil tS a tS Sil Yes Yes 2 Sil-tS, tS-Sil YES 2 Yes YES 2
_cha_n_cha... F:\... 1.... tS a n tS a tS Yes Yes 3 tS-a, n-tS, a-tS YES 3 Yes YES 3
_che_n_che... F:\... 2.... tS e n tS e tS Yes Yes 2 tS-e, e-tS YES 2 Yes YES 2
_chi_n_chi_c... F:\... 2.... tS i n tS i tS Yes Yes 2 tS-i, i-tS YES 2 Yes YES 2
_cho_n_cho... F:\... 1.... tS o n tS o tS Yes Yes 2 tS-o, o-tS YES 2 Yes YES 2
_chu_n_chu... F:\... 2.... tS M n tS M tS Yes Yes 2 tS-M, M-tS YES 2 Yes YES 2
_da_d.trans F:\... 1.... Sil d a d Sil Yes Yes 2 Sil-d, d-Sil YES 2 Yes YES 2
_da_n_da_d.... F:\... 2.... d a n d a d Yes Yes 3 d-a, n-d, a-d YES 3 Yes YES 3
_de_n_de_d.... F:\... 2.... d e n d e d e d Yes Yes 2 d-e, e-d YES 2 Yes YES 2
_di_n_di_d.tr... F:\... 2.... d' i n d' i d' i d' Yes Yes 2 d'-i, i-d' YES 2 Yes YES 2
_di_n_di_d+.t... F:\... 2.... d' i n d' i d Yes Yes 1 i-d YES 1 Yes YES 1
_do_n_do_d.... F:\... 1.... d o n d o d o d Yes Yes 2 d-o, o-d YES 2 Yes YES 2
_du_n_du_d.... F:\... 2.... d M n d M d Yes Yes 2 d-M, M-d YES 2 Yes YES 2
_dya_dy.trans F:\... 1.... Sil d' a d' Sil Yes Yes 2 Sil-d', d'-Sil YES 2 Yes YES 2
_dya_n_dya_... F:\... 2.... d' a n d' a d' Yes Yes 3 d'-a, n-d', a-d' YES 3 Yes YES 3
_dye_n_dye_... F:\... 2.... d' e n d' e d' Yes Yes 2 d'-e, e-d' YES 2 Yes YES 2
_dyo_n_dyo_... F:\... 2.... d' o n d' o d' Yes Yes 2 d'-o, o-d' YES 2 Yes YES 2
_dyu_n_dyu_... F:\... 1.... d' M n d' M d' Yes Yes 2 d'-M, M-d' YES 2 Yes YES 2
Step 1: Phoneme-Level Segmentation (Stationaries And Articulation Compounds)
Step 2: Estimation Of EpR And Adding To Database Of Stationaries
Step 3: Optimization Of EpR Guides From EpR Of Stationaries In Database
Step 4: Articulation-Level Segmentation, Estimation Of EpR (Using EpR Guides) And Adding To Database Of Articulations
Edit Transcription
Segment Phonemes
View/Edit Segmentation
Edit Wave Data
Metro-segment Phonemes
Metro tempo:
120
BPM
Add Stationaries To Database
Remove Stationaries From Database
Manually Adjust EpR Of Stationaries (Exit, Adjust And Return To Dialog)
Optimize EpR Guides
Segment Articulations
View/Edit Segmentation
Add Articulations To Database
Remove Articulations From Database
Abort Runing Task(s)
Exit

Select them all and Segment Phonemes. If it says error, that's normal. It'll generate the base markers. After which you'd open them in view and fix them accordingly. Be careful. If you edit the Transcript you will have to redo your segmentation. **Not being careful will lead you to need to redo them and will force it to remove all the just added samples.**

If it says Error/0 under Stationary or Articulation the transcript is broken. Simply go in and correct the phonetics. When it's showing you less articulations than you know you should have, there's also something wrong with the transcript.

When it says **BAD** next to anything, your markers are most likely overlapping or the phonetic is not in the dictionary. You will have to redo your Segments. Sometimes it doesn't matter… If there's a 0 and nothing else next duration, there's no sample!

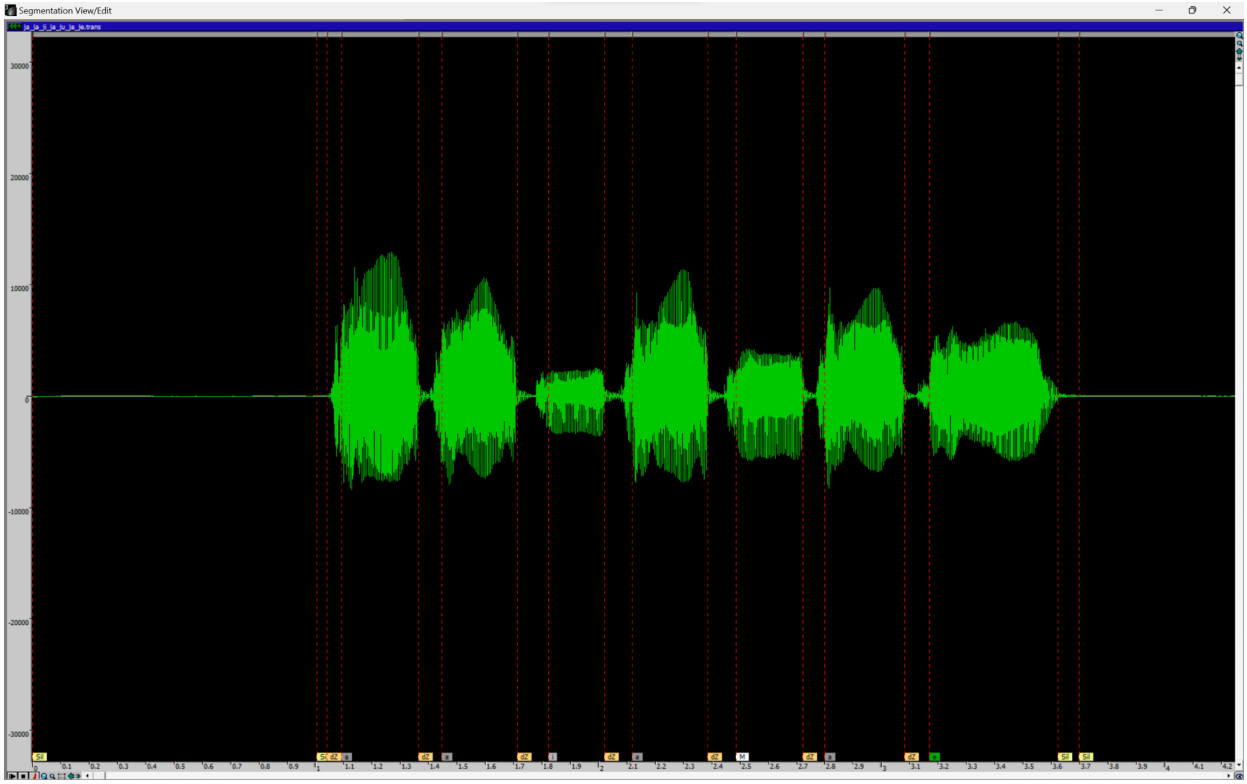
If you double click on the transcript, stationary, or articulation their respective viewers will open. Double click on the file name will bring up your external audio viewer. You can also just right click something and the same options will appear. When you have the chance go into File and set up an external audio editor. It can be Audacity, anything is fine.

Step 1 - Editing Segmentations

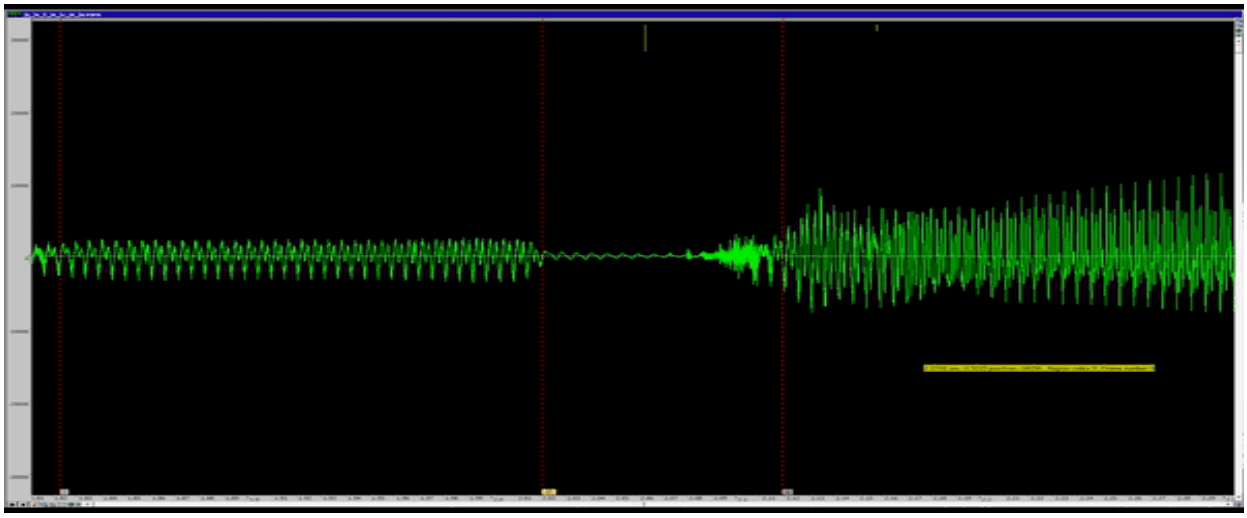
You can **L Ctrl or LShift + Click to select multiple entries to edit.** The max allowed by the tool is **6**.

Click View/Edit Segmentation

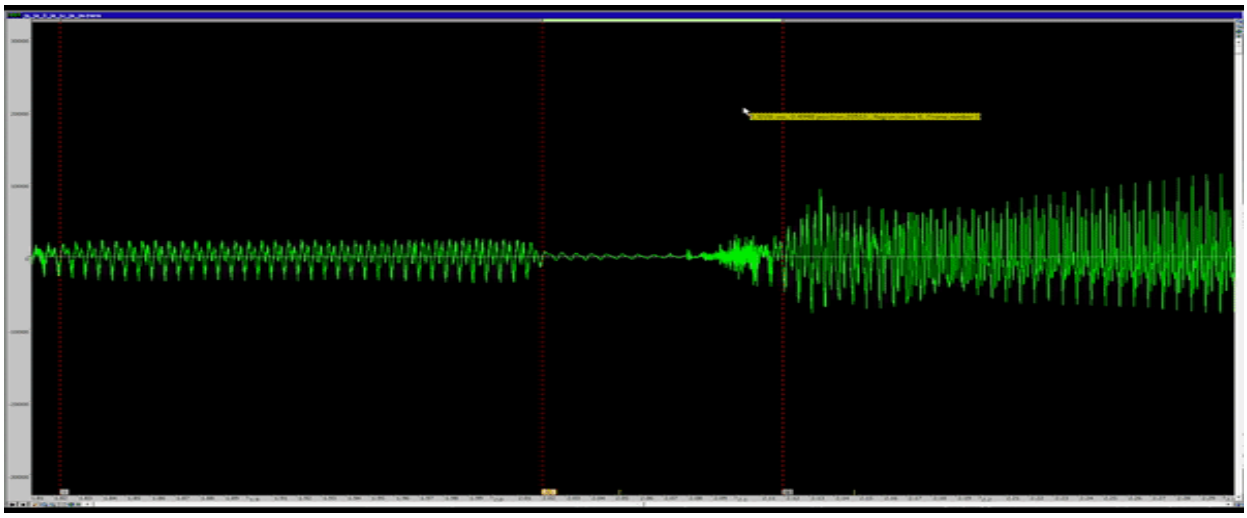
Example is for: ja_ja_ji_ja_ju_ja_je



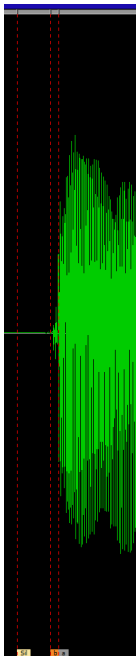
You'll click **the top grayed out bars** to allow selection of the Phoneme Segment. This allows you to **select the red bars** to edit segments of the phonemes



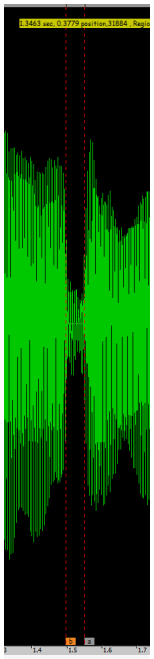
Then you'll be able to **select the red lines** and move them as you like.



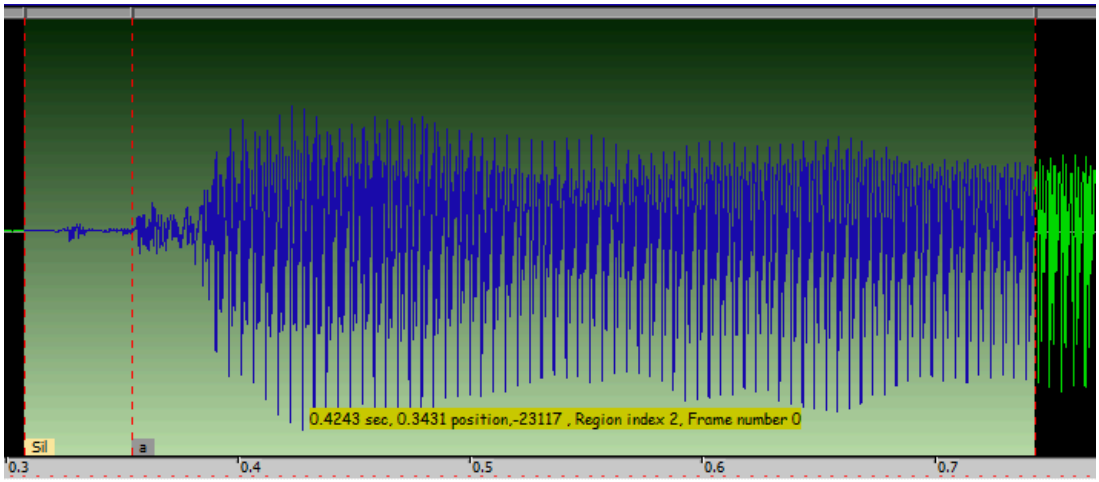
Placement is very important, this will minimize the amount of edits you'll need to make in the Articulations menu later.
For Sil C V you'll want to place the marker at the start of the Phoneme,



For a C V that is after a V you'll place the consonant at the end of the previous note & place the Vowel after the Consonant in the sample.



For Sil V sound in the marker is fine. (not nasals, or liquids).



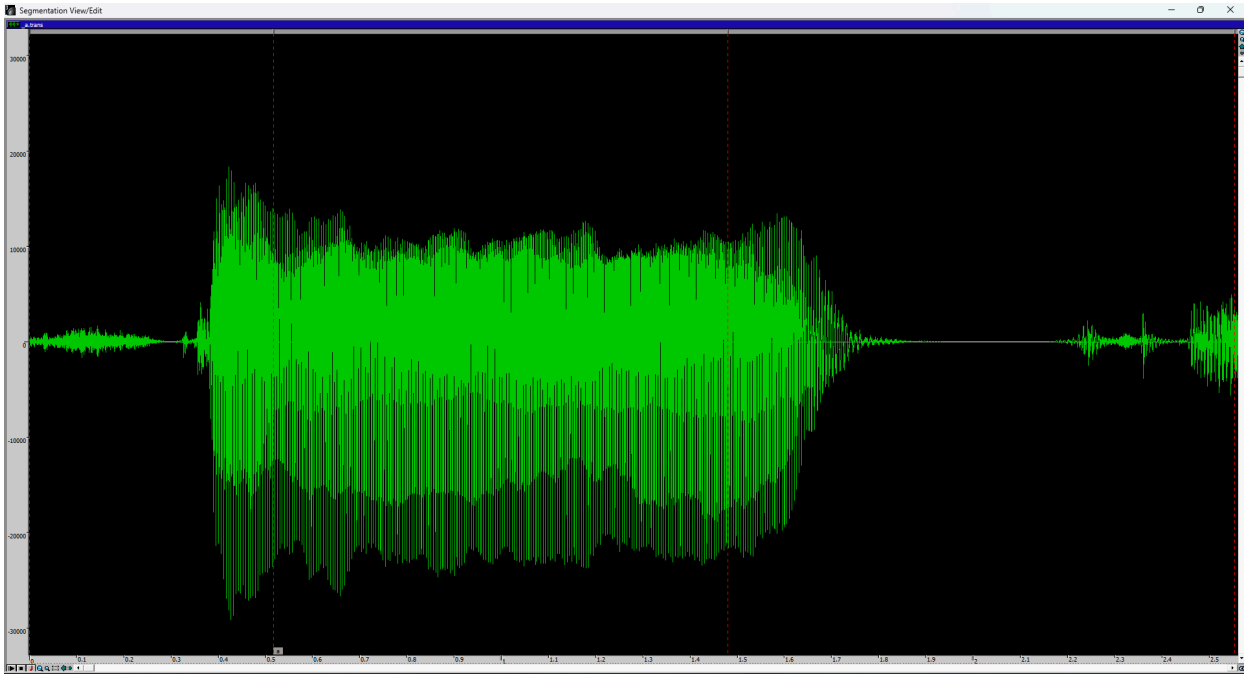
(Confused? Look at the example bank in **resources**)
To listen to samples simply press **Space**.
To listen to a select part of a sample **hold Shift and Drag**. To deselect, **Hold Shift and Click**.
To select a sample in a marker **hold Shift and Click the marker**.
If you're having issues seeing samples. **Use the side scrollers (The green arrows) to zoom in by dragging up and down or left and right.** (All menus will have them)
Close the UI **(x)** and a **YES** will appear next to the sample. Meaning its markers have been edited.
When editing markers, if you don't see a phonetic on the bottom of the marker, **the phonetic is not in the dictionary.**
(These rules apply to all menus in and out of the autoseg)

Step 2 - Stationaries

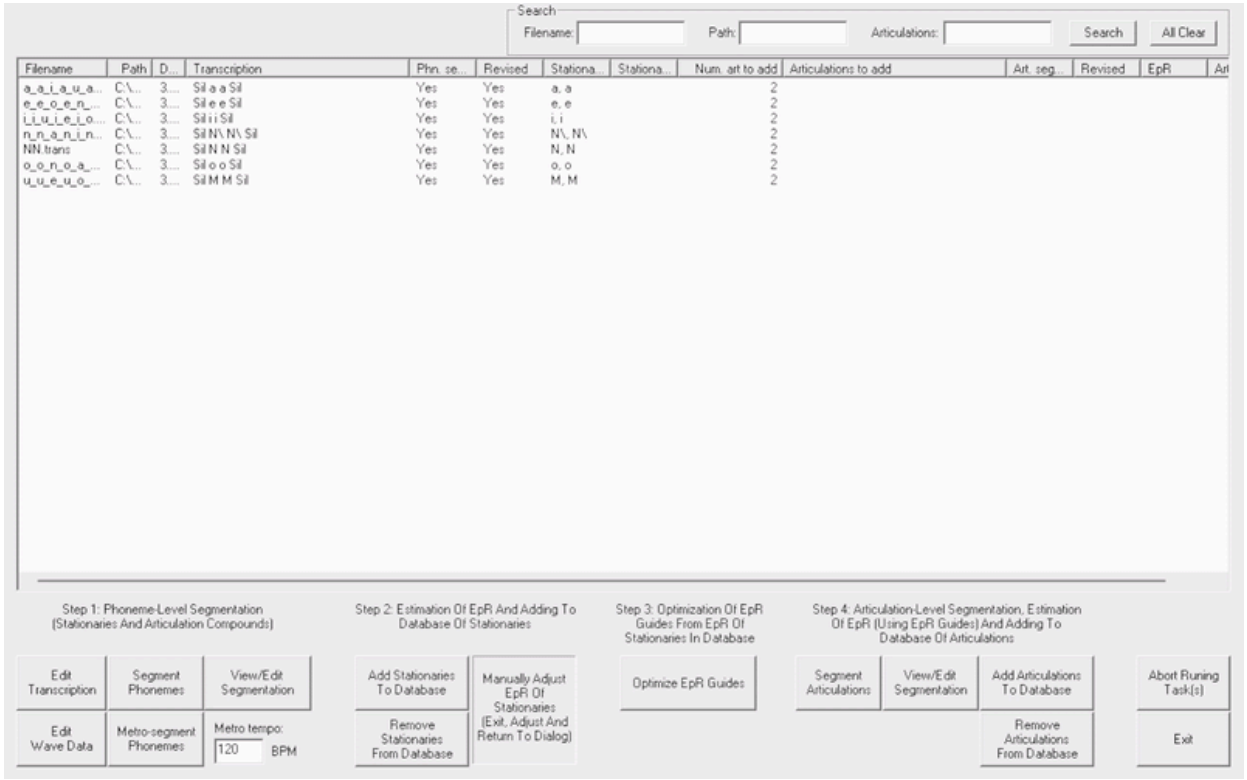
To add stationeries is as follows
1: Segment Phonemes
2: View/Edit Segmentation
3: Add Stationaries to Database

Select the Sample(s) you want to Segment and click **View/Edit Segmentation.**
This may take several seconds, but once it's done click the segment you want to edit and select **View/Edit Segmentation.**

Select a piece of your C/V Sample(s) long or short. **Make sure the marker is not in the beginning or ending of the C/V sample.** In the middle or some place where the stabilization is clean. The longer the better.



Then go ahead and add it into the bank.



***Sadly this will not be enough for a test bank. Add in a Sil C/V or a C/V Sil of the same phonetic to test your samples.**

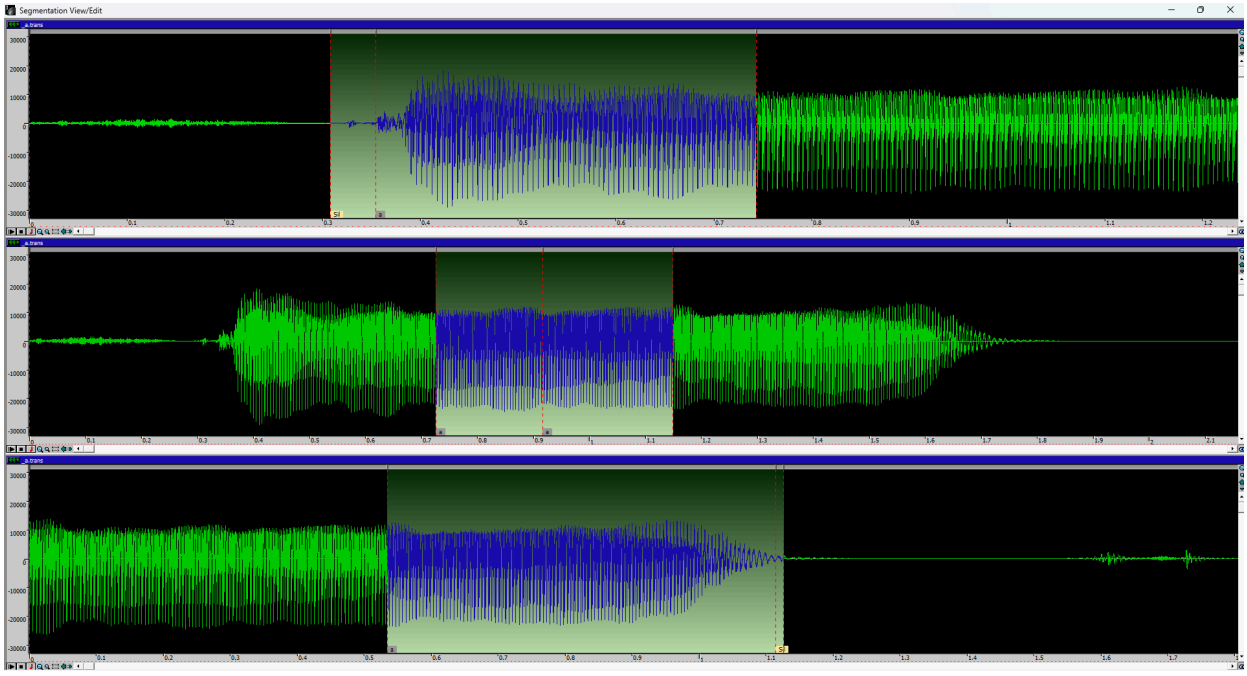
YOU CANNOT SKIP STEP 2, AND YOU ONLY DO THIS FOR STATIONARIES.

Step 3 - Articulations

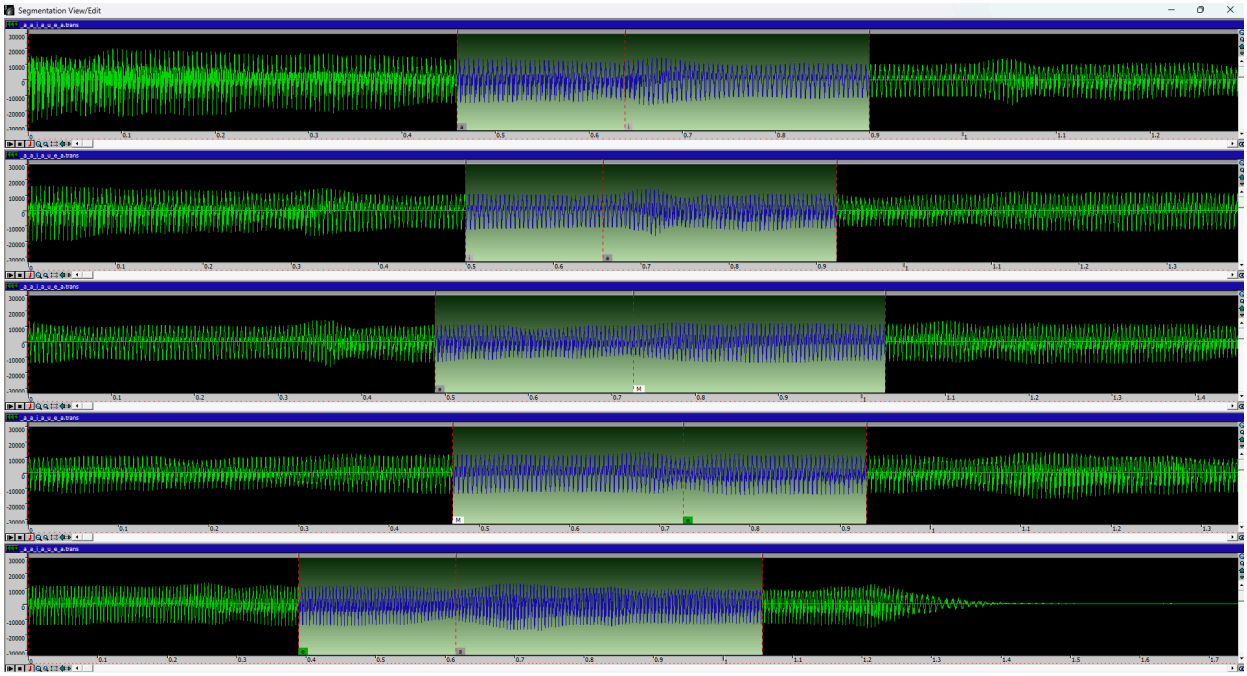
To add Articulation is as follows
1: Segment Phonemes
2: View/Edit Segmentation
3: Segment Articulations
4: View/Edit Articulations
5: Add Articulations to Database
YOU CANNOT SKIP STEP 2 OR 4.

Select the Sample(s) you want to Segment and click **Segment Articulations**
This may take several seconds, but once it's done click the segment you want to edit and select **View/Edit Segmentation.** You can only do **1** at a time.
***How many brackets you make determines the number of Articulations. C-V V-C, and V-C C-V both fall under this!**

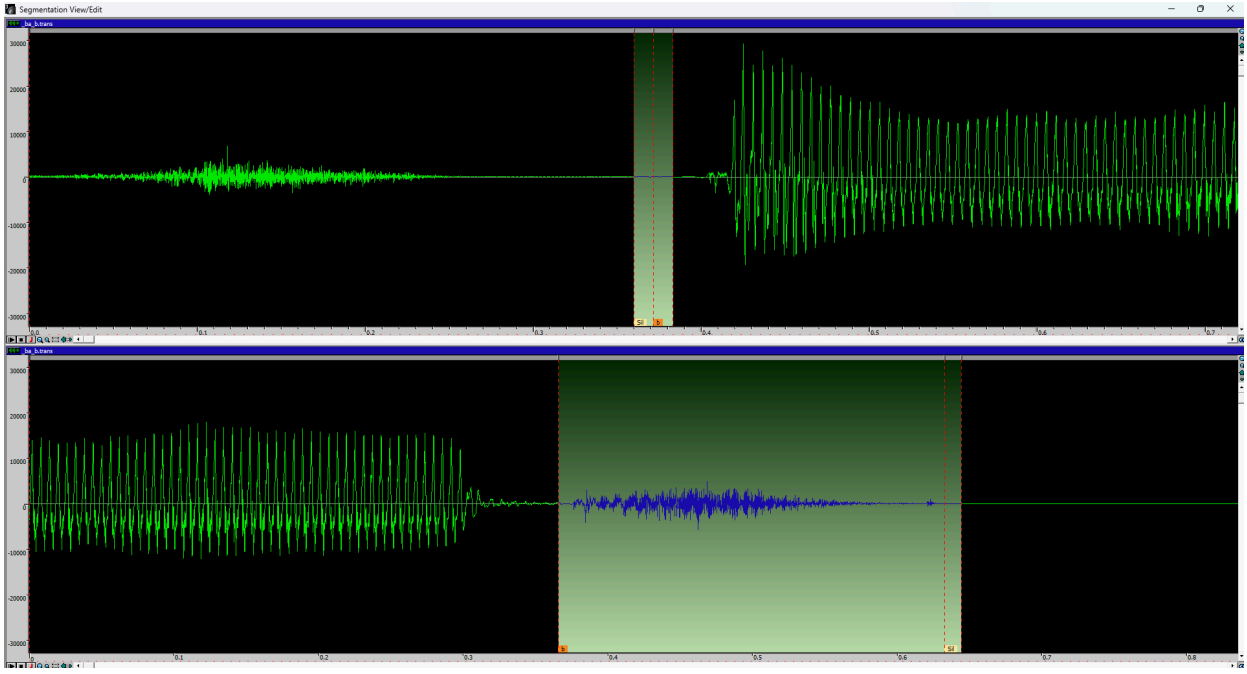
Example is for: [- a] [a a] [a -]



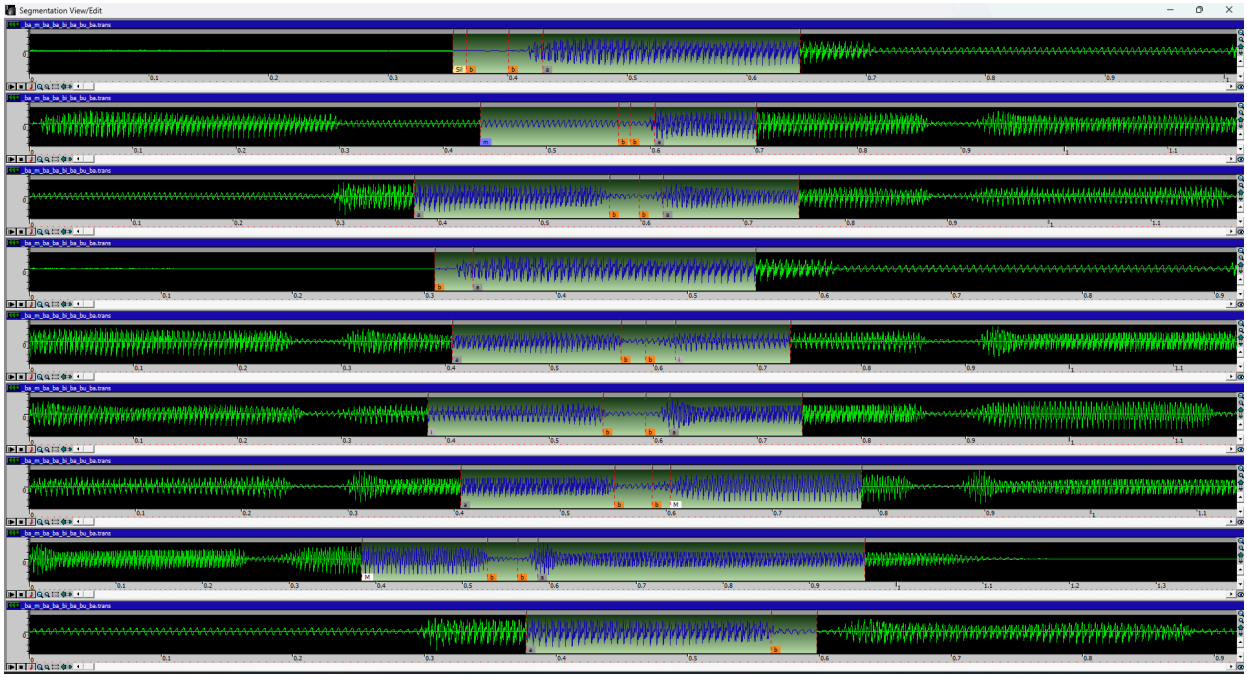
Example is for: [a i][i a][a M][M e][e a]



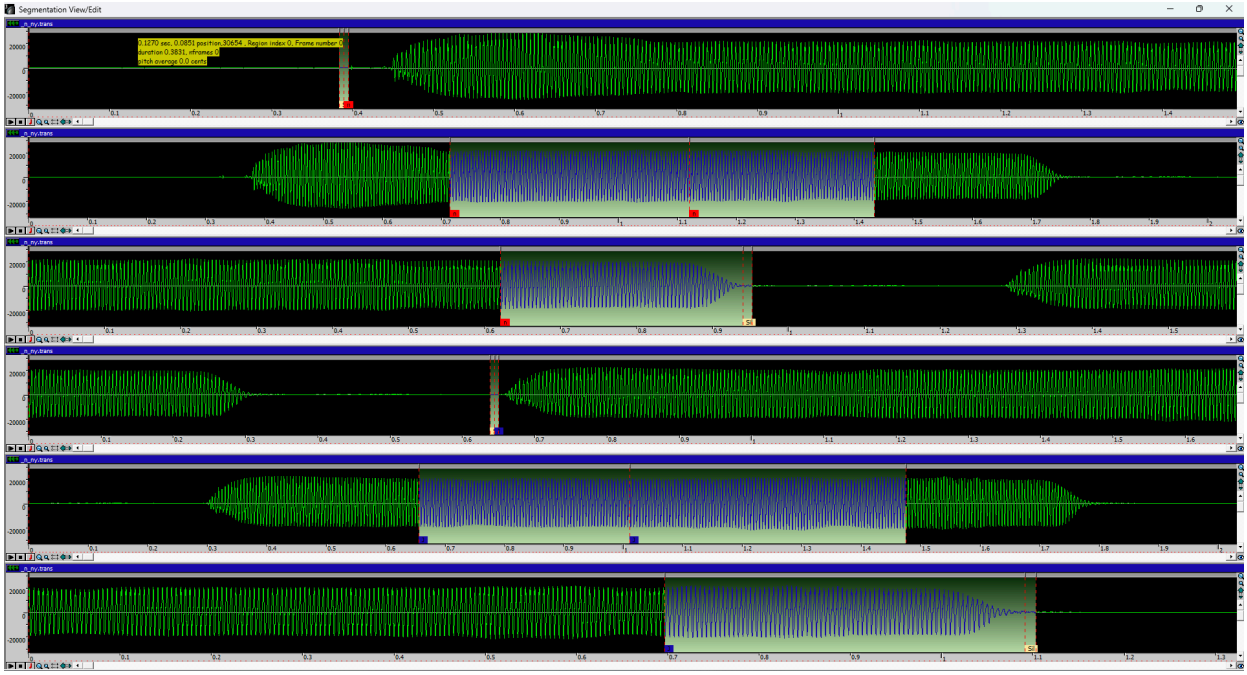
Example is for: [Sil b][b Sil]



Example is for: _ba_m_ba_ba_bi_ba_bu_ba (V-C C-V/C-V V-C)



Example is for: [Sil n][Sil J]



The process is the same as editing **Phoneme Segmentation**. but this time you'll just want to make sure everything is aligned correctly.

You can add the files into the database by clicking **Add Articulations To Database** Depending on the amount you've selected this can take Seconds to Several Minutes.

***ADVICE:** To get a better sounding voicebank, barely add in any of the vowels when doing articulations. Just enough of it.

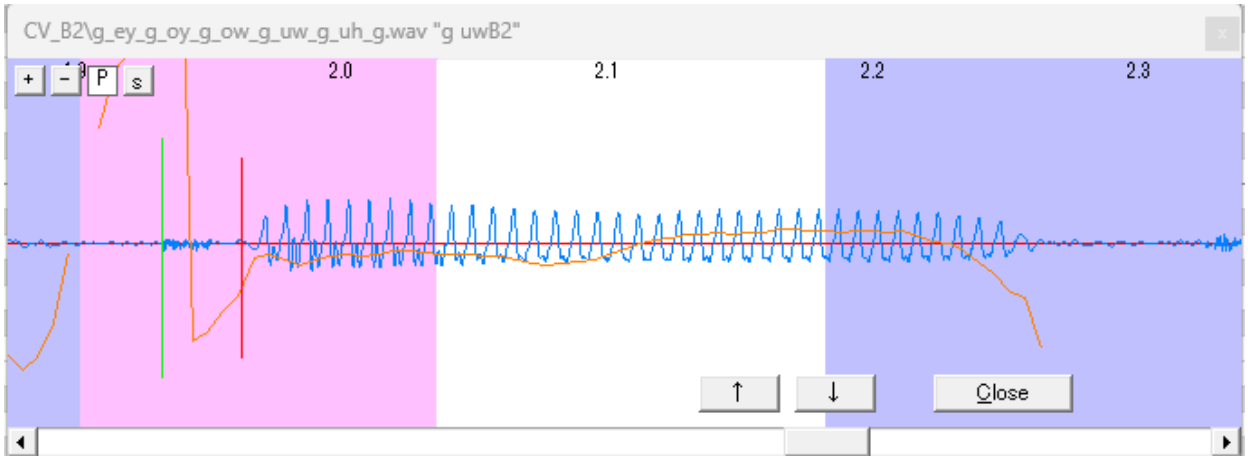
UTAU to VOCALOID (The Oto'er Guide to The Galaxy)

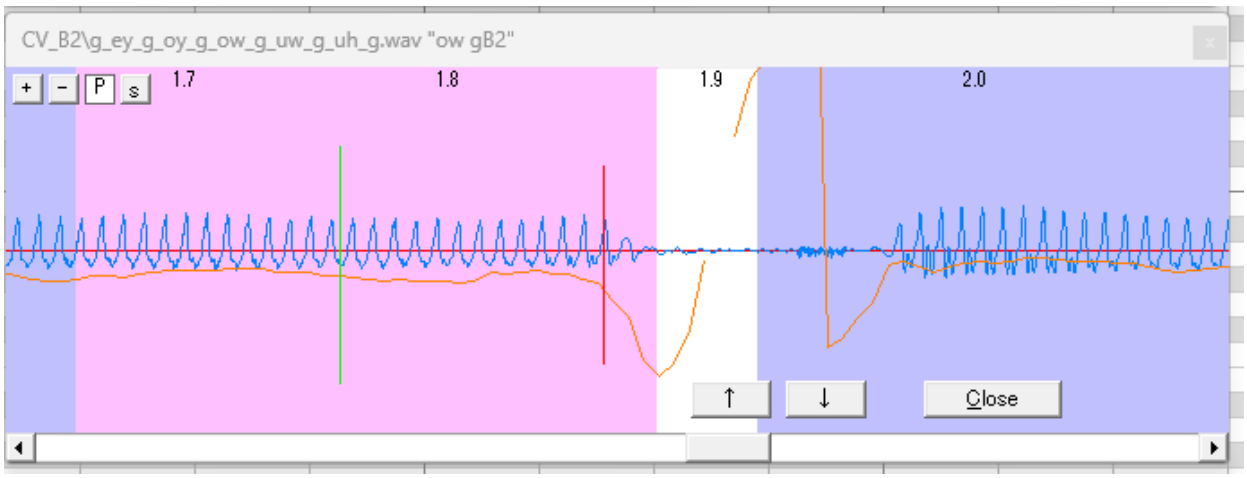
It's no secret how deep the learning curve is for vocal synth development, but with these visual guides I'll make it look super easy.

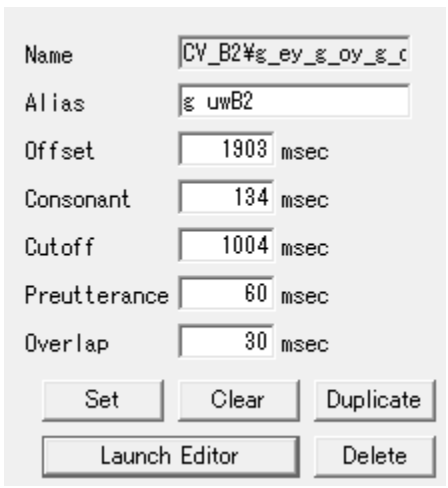
In UTAU when you oto you start out with a blank slate. You first look at the **name** of the sample you're editing, Type in the **alias** for the part of the sample we want to isolate, Use both **offsets** to isolate the section, then use the **consonant** for the piece of the sample we always want to play. The **Preutterance** is where our vowel starts, and the **overlap** is where our crossfade starts!

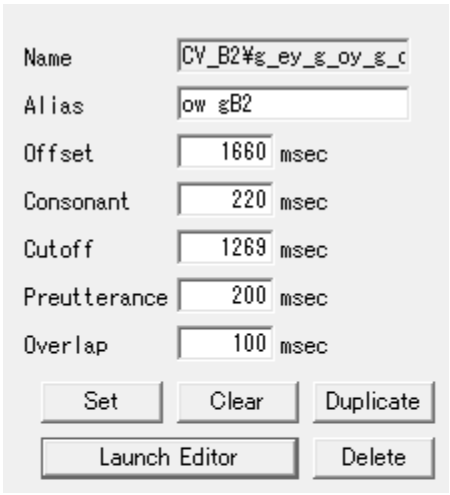
Like how I talked about transcription before, it's telling the devkit what pieces you want isolated and sectioned off into the voicebank to be crossfaded together to make our bigger sound. The only major difference is, **you have to section out what is where first.** If you have a 'babi', You need to line up each line with each phonetic. If there's a 'b' where does said 'b' start and end. Same thing for 'a', the next 'b' and so on.

Example is for: CVVC

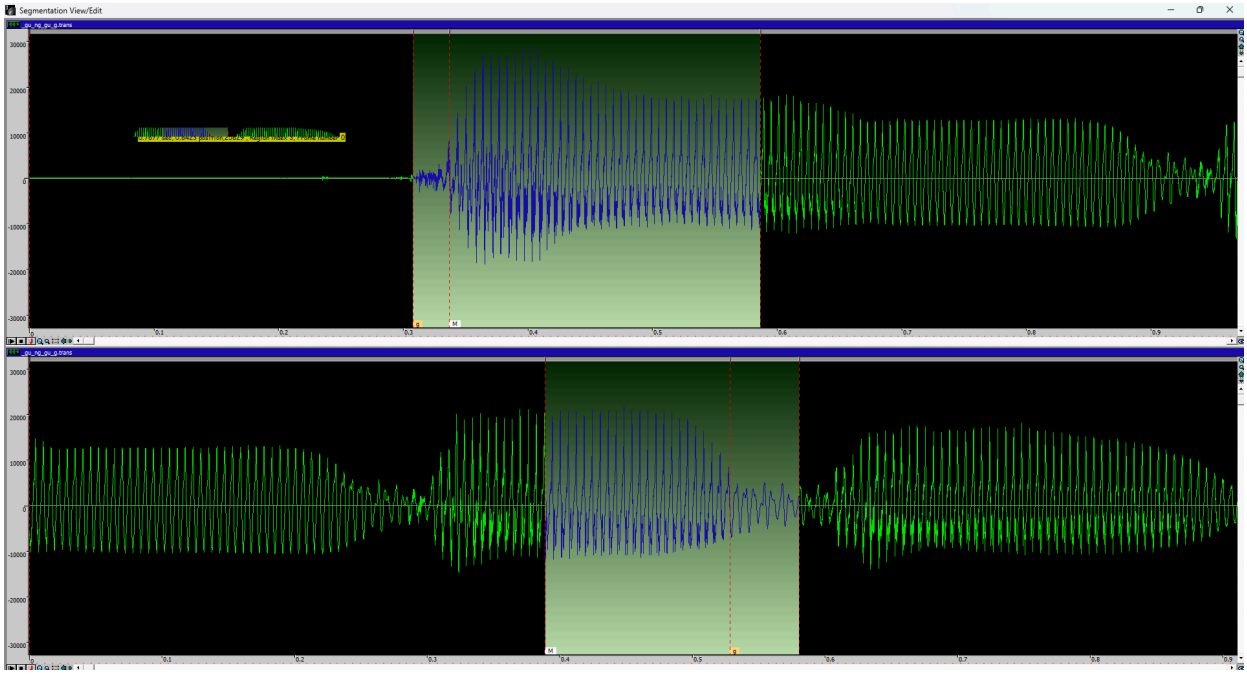








These may be for English but it’s about the same. Here’s the same difference in Vocaloid.



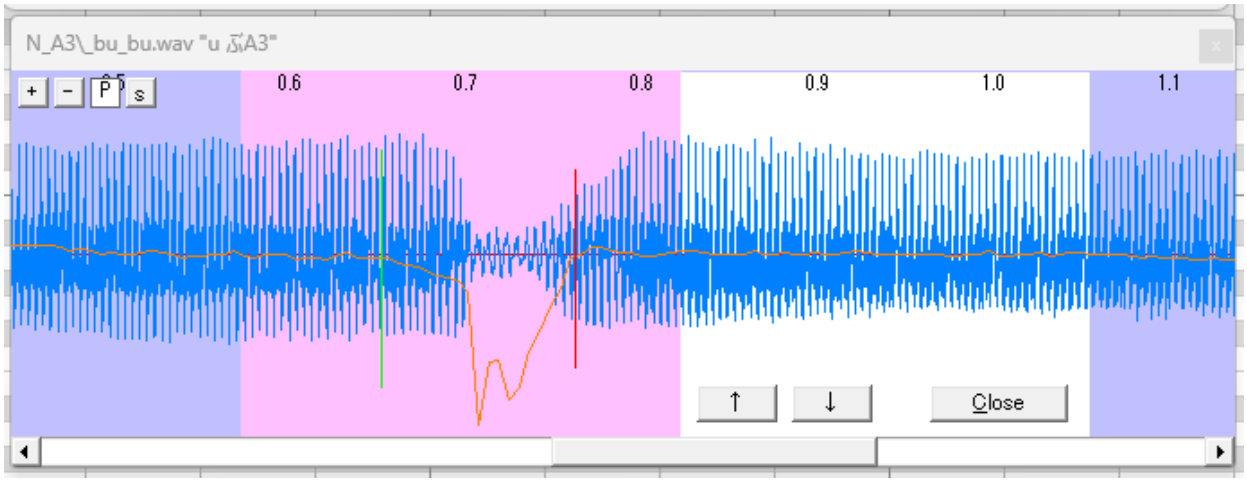
The connecting VC almost never has the C directly in it, but what leads up to that connecting sound.

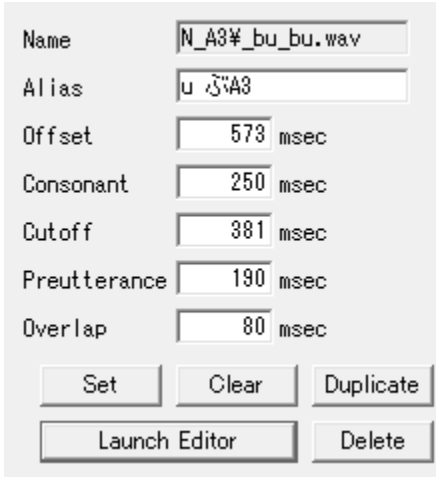
You can see that your **overlap** is usually the first line, and your **Preutterance** is your second one. What we'll call line 0 (The line before your **overlap** which doesn't get included if it's not in the marker) is your 1st **offset**. Your **consonant** is where your 3rd line ends. Making that point the 2nd **offset**. When it comes to configuration, **you want to treat Vocaloids like they're Arpasing banks.** They're in theory built the same way, or really lacking the stationary. You don't need multiple openings, and multiple closes however! Only 1 of everything for each pitch. You can even live with doing 1 line of [Sil C]/[Sil k] samples because they're just silence. You also don't exactly need more than 1 pitch of [C sil]/[k Sil], because of their functionality.

And think of it like this, the segment has a C on the end as a connection. You don't want that sound in the articulation, but rather what leads up to it!

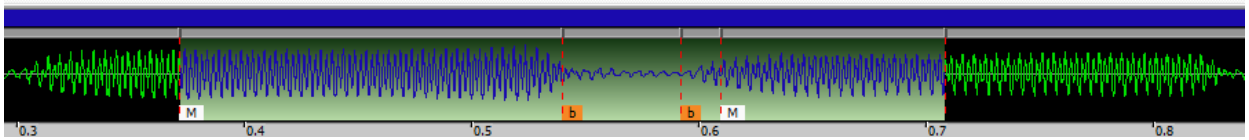
Now for a VCV string

Example is for: VCV





This too changes slightly for Vocaloid.

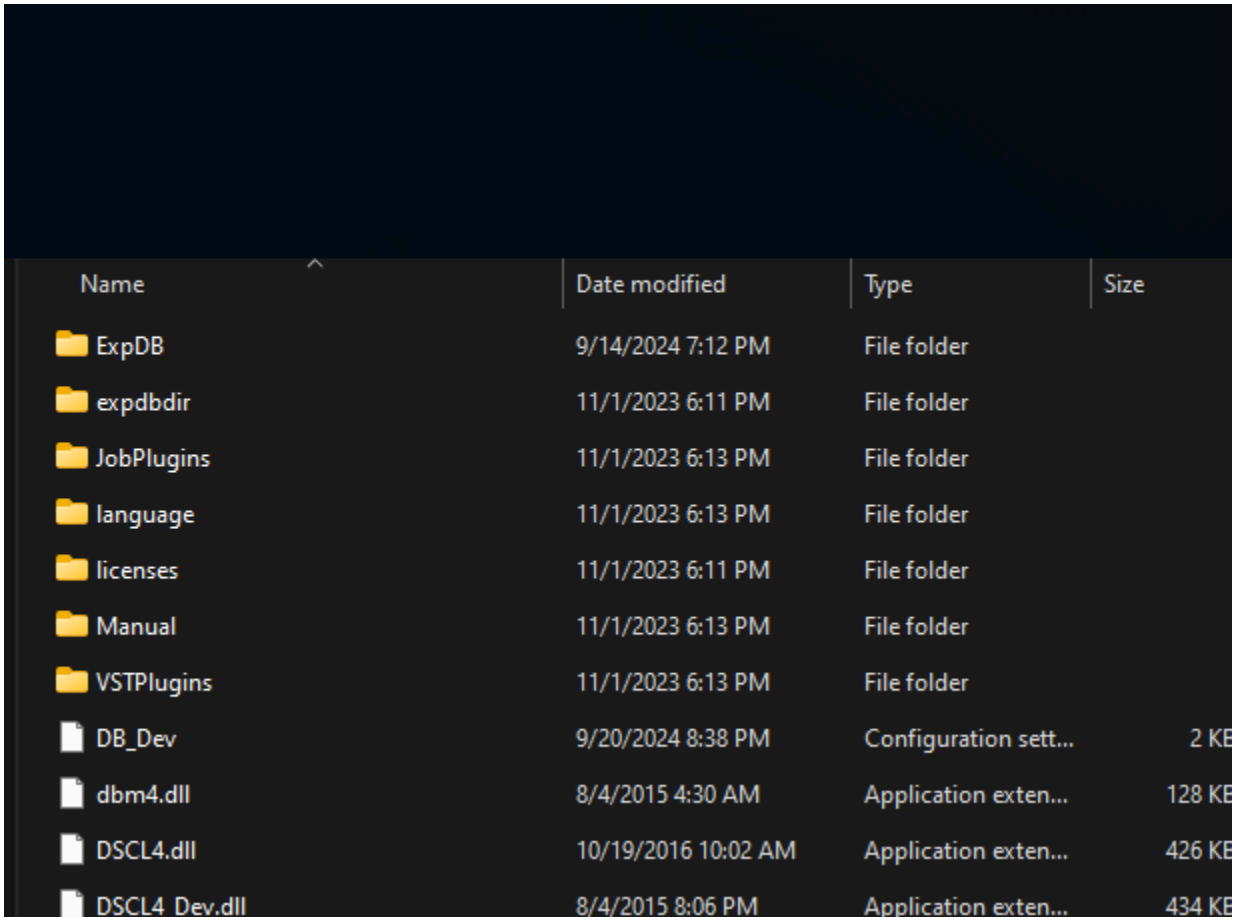


You have both your **offsets**. A more controlled **overlap** and your With your **Preutterance** being the second vowel, or just the vowel after the **overlap** in general. The **consonant** isn't really a thing in Vocaloid the way it is in UTAU, because of the way things are segmented. However as you can see there's an overall gist. Here you're (After segmenting) correcting your isolations. Isolating the connecting Vowel or silence, then isolating the lead up the C, and then the C itself. Afterward you isolate the V, and you have your VCV configuration. This example above isn't the same for all phonetics, but a general gist. **Think of it more as V C, C V.**

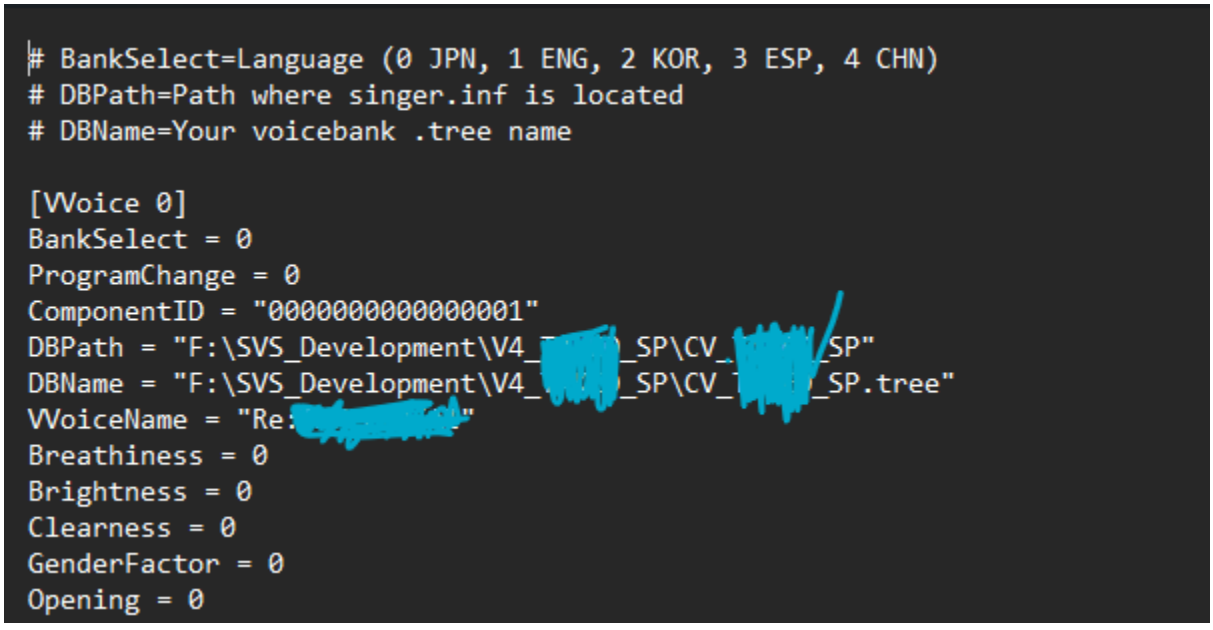
VOCALOID DEV LOAD IN

Install Vocaloid Dev Editor and follow the instructions.

With the Dev_DB file, Link your singer like this (Make sure you use Notepad++)



Name	Date modified	Type	Size
ExpDB	9/14/2024 7:12 PM	File folder	
expdbdir	11/1/2023 6:11 PM	File folder	
JobPlugins	11/1/2023 6:13 PM	File folder	
language	11/1/2023 6:13 PM	File folder	
licenses	11/1/2023 6:11 PM	File folder	
Manual	11/1/2023 6:13 PM	File folder	
VSTPlugins	11/1/2023 6:13 PM	File folder	
DB_Dev	9/20/2024 8:38 PM	Configuration sett...	2 KB
dbm4.dll	8/4/2015 4:30 AM	Application exten...	128 KB
DSCL4.dll	10/19/2016 10:02 AM	Application exten...	426 KB
DSCL4_Dev.dll	8/4/2015 8:06 PM	Application exten...	434 KB



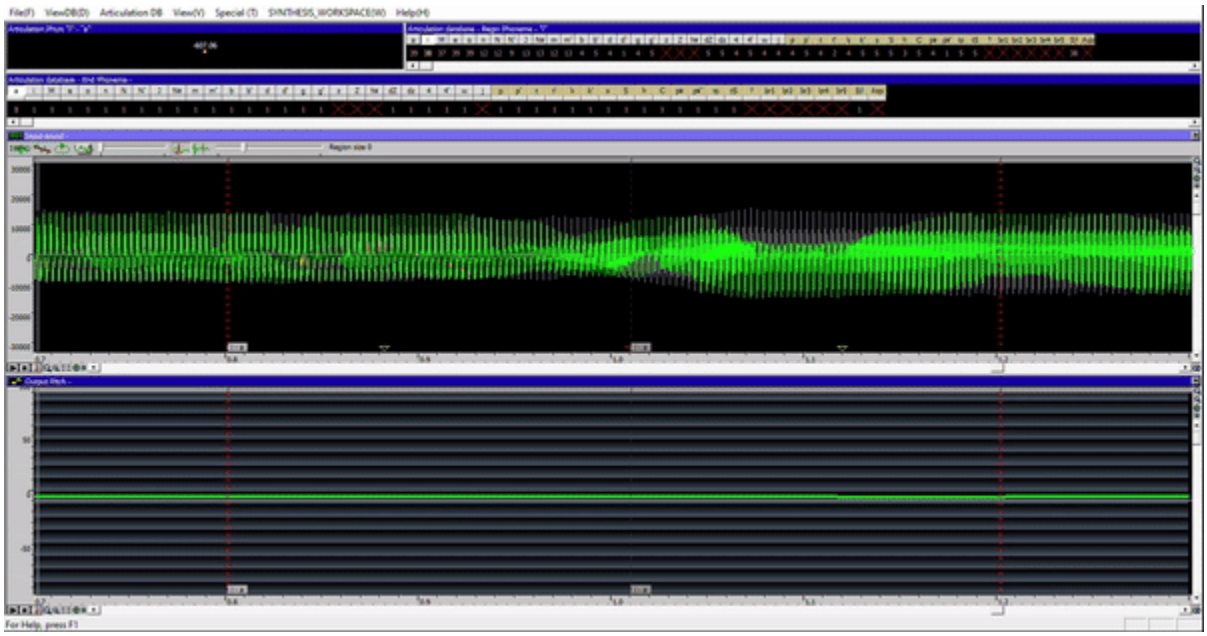
```
# BankSelect=Language (0 JPN, 1 ENG, 2 KOR, 3 ESP, 4 CHN)
# DBPath=Path where singer.inf is located
# DBName=Your voicebank .tree name

[VVoice 0]
BankSelect = 0
ProgramChange = 0
ComponentID = "0000000000000001"
DBPath = "F:\SVS_Development\V4_[redacted]_SP\CV_[redacted]_SP"
DBName = "F:\SVS_Development\V4_[redacted]_SP\CV_[redacted]_SP.tree"
VVoiceName = "Re:[redacted]"
Breathiness = 0
Brightness = 0
Clearness = 0
GenderFactor = 0
Opening = 0
```

Please remember as you update, you will need to restart the editor. Hell if you don't want to pack your voicebank. Just keep it in the dev editor. It'll work just fine!! Be aware that Vocaloids Octaves are off one. C4 is C5, and so on.

Your paths should be the same length, the only difference being .tree at the end of the second.

F0 Editor/OutputAttPitch

Comes in handy for those that wish to pitch correct their data. Works similarly to UTAU .frq and World .dio files. It's quite literally Vocaloid's .frq editor so to speak. (it would be Output Pitch you'd edit)



For **Articulation** data, the window will be toggled on by itself.
For **Stationaries**, users can enable it themselves via: **View > OutputAttPitch**
Holding CTRL while drawing on the editor's window allows you to draw your own pitch points
Holding CTRL and D at the same time while drawing allows you to delete pitch points.
Once you're done editing, save your progress via: **Stationary/ArticulationDB > DB Update**

-Daigasso

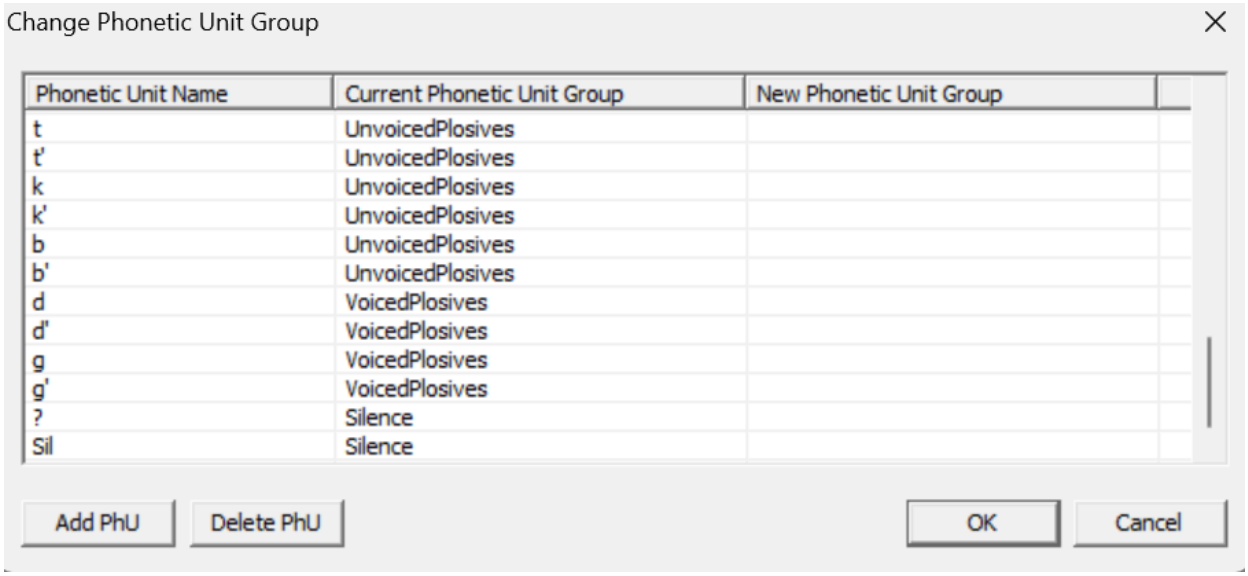
***ADVICE:** If you're getting weird FRQ noise, just add a line to connect the samples.

ADDING/REMOVING PHONEMES

***This closes the DB (Specifically the bank) after use, make sure you have stuff saved beforehand. Simply reopen it when you're done.**

You can add additional phonemes into the editor, but you'll have to manually input them. This can be useful to add additional support for another language, or special extras. Be wary of what you add, because it will only work in editor with manual phonetic input depending on the language.
The DBTool's segmentation will not recognize/refuse phonetics in the transcript that are not in the dictionary.

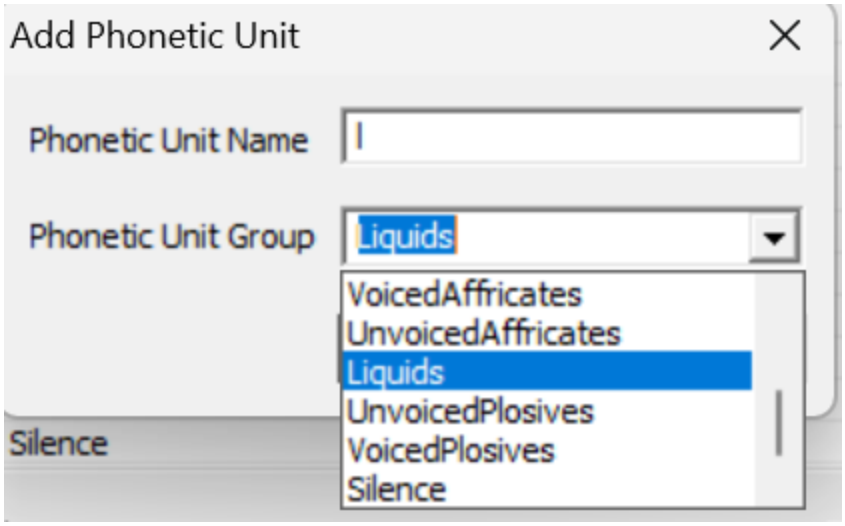
Special -> Change Phonetic Unit Group



If we want to add **L** to the system, we'll click Add **PhU.**

Then you'll enter the **Phonetic Unit Name** and the **Phonetic Unit Group**

L is our phoneme we're wanting to add, so we'll need to select the dropdown and select what group it belongs to.



Alternatively, you can make edits to the dictionary text file beforehand. Use others as examples
I recommend you add br samples as whatever is voiced. (Not just br, i mean br1, br2, etc! They will need an opening and ending! The opening should have no sound, and the ending [br1 Sil] is where the breath comes from.) **You can add vocal fry phonetics too, no guarantee it'll work tho.**
This is one way to create EVEC

PACK TO DDB

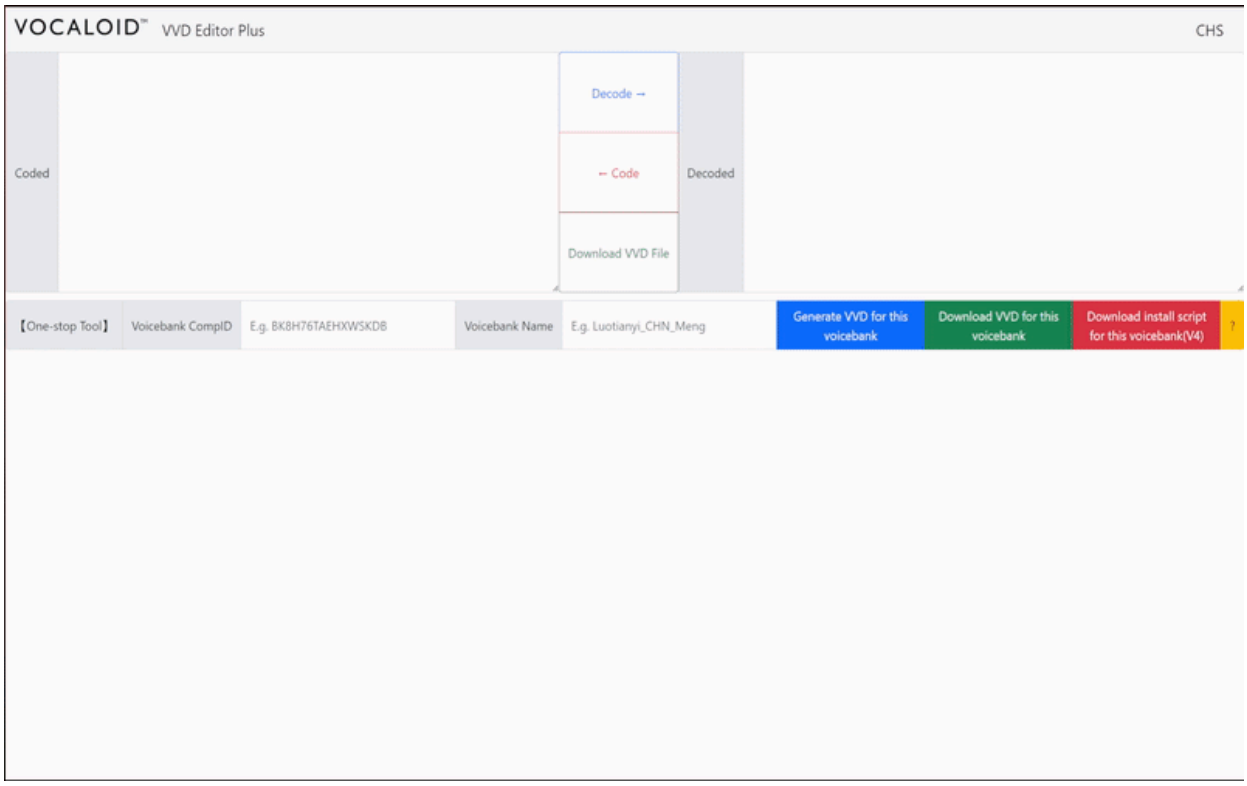
Recommended to do this after you've finished your Voicebank, this can sometimes make your bank's storage size-able.

This will allow your Voicebank to be used on Vocaloid V3 & Up

Pre Prerequisite:

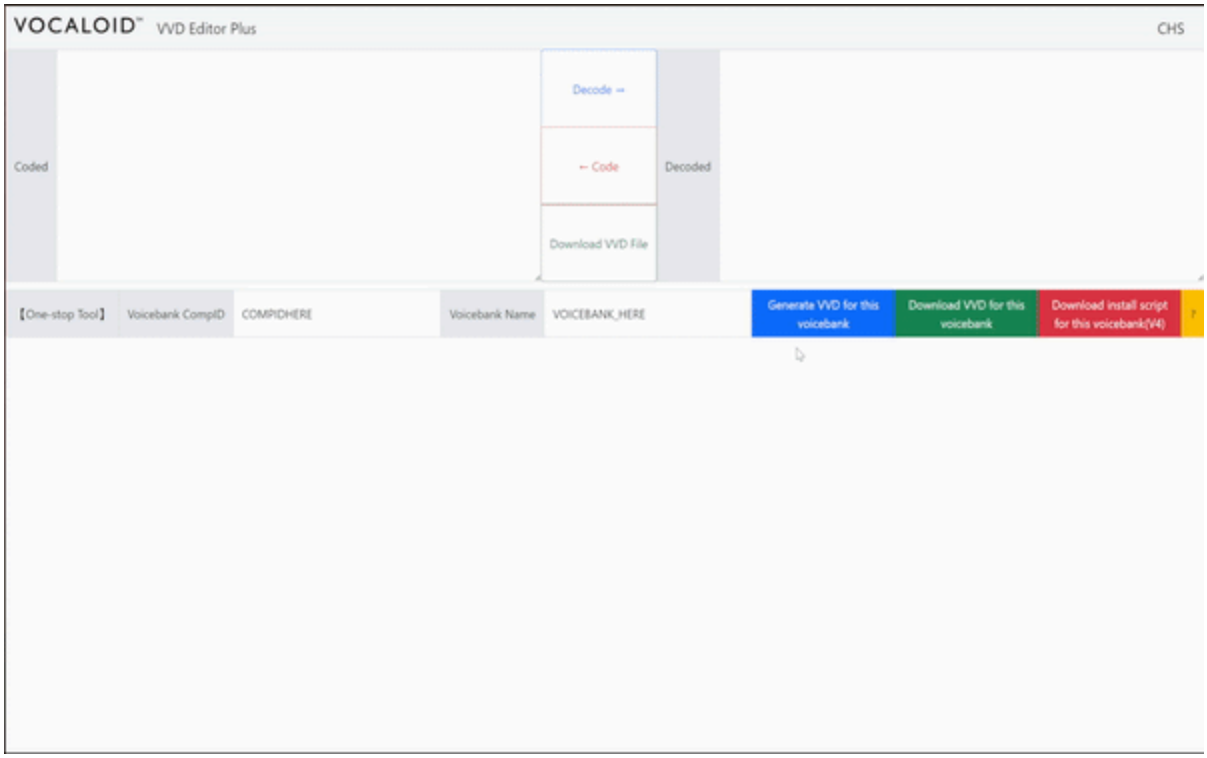
- Python 3.10.0
- Pyyaml (do through pip install, inside terminal)
- Download and extract the DDB_Packer.zip. "Found in **Tools & Software"**

You'll need to generate a VVD, Install & Uninstall script for your Voicebank. You can do that Here.
Click the **?** to view CompID resources, pick a CompID for the Voicebank.



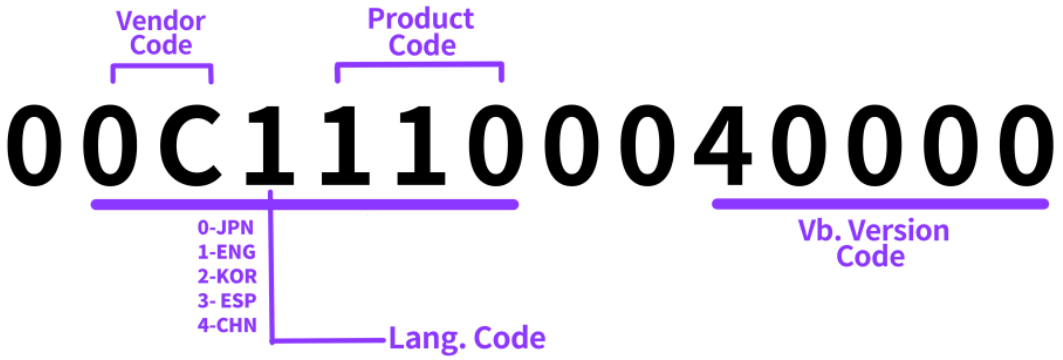
You'll then take that CompID place it inside of the **Voicebank CompID** Section. and provide the **Voicebank Name**. Click **Generate VVD**

Change Format from 4.0.0.0 to 3.0.0.0
Change Growl from 1 to 0
Click Code
Click **Download VVD** & **Download Install Script**



Another way to generate a CompID is with the Keygen found in resources

Decoded structure :
[0][DRP(6 digits)][00][Version Code(5 digits)]



The DRP has six digits, of which the first two are the vendor code, the third is the language code, and the last three are the product code.

Here's a quick guide made by Daigasso!

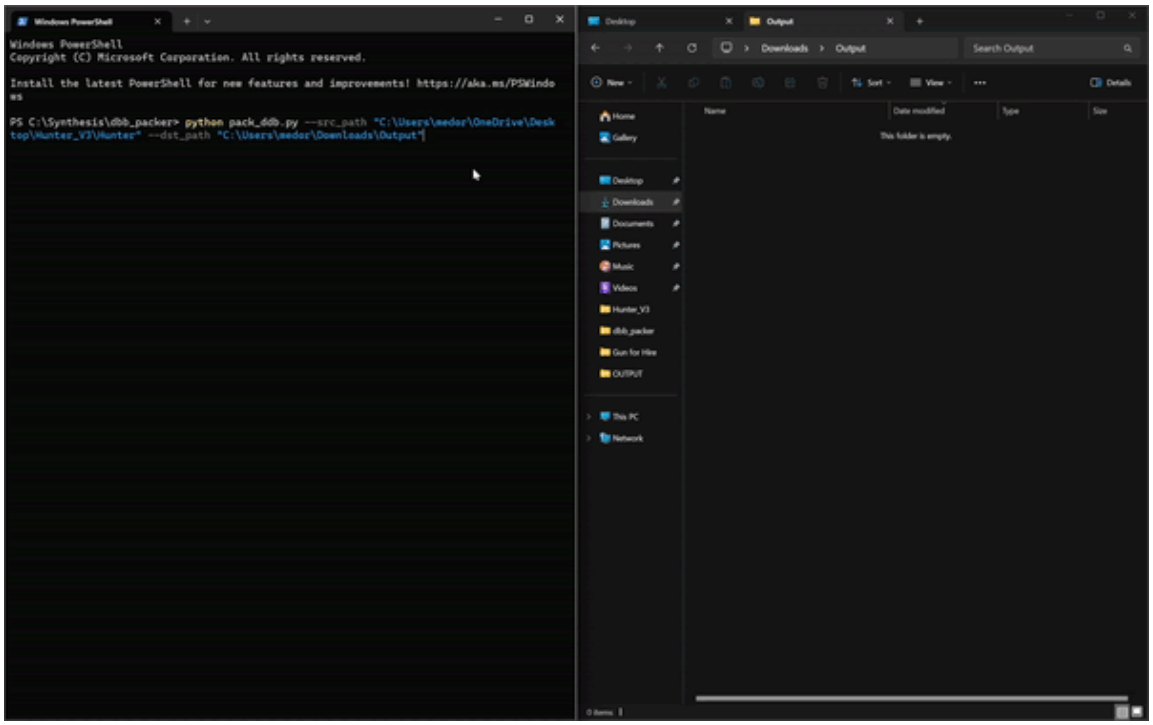
***Feel free to skip this step if you're gonna use the GUI found in resources.**

Go to the folder where you extracted the DDB Packer. **Right Click** & select **Open In Terminal**
A terminal will open, this is where you'll prepare the DB to be packed.

Now we need to pack the Voicebank. Create a folder and name it the name you gave the Voicebank, inside that folder. create another & name it the CompID.
rename the VVD to the name of the Voice, move it and the install/uninstall scripts into the newly created folder.

Navigate back to the Terminal & type the following:

python pack_ddb.py --src_path "the folder where your singer.inf is located" --dst_path "the folder path we just created"


Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.
Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows
PS C:\Synthesis\ddb_packer> python pack_ddb.py --src_path "C:\Users\eedor\OneDrive\Desktop\Hunter_V3\Hunter" --dst_path "C:\Users\eedor\Downloads\Output"
Output
Downloads > Output
Search Output
This folder is empty.

Once all the files have been created, you'll move the folder into the **VoiceDB** folder and run the install script as an Administrator.

For V5 & V6 run **Vocal Register/Vocareg**

*For some reason some mixin's don't work with CVVC. There's another way to do it, but we're not entirely sure. **(To circumvent this just do one VCV segment ([V C V]) during development. Doesn't matter the language.)**

Add growl to DDB with Python

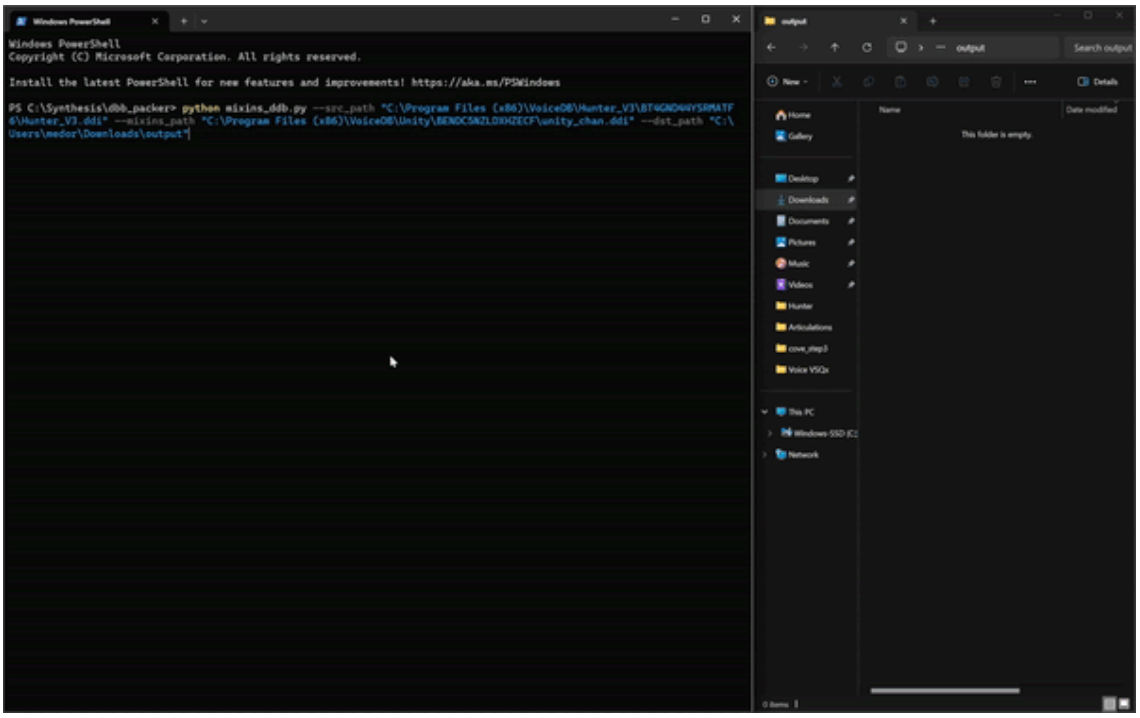
You must have a packed DDB for this to work

You'll need to acquire a Voicebank that already has a growl in it. V4Flower, VY1, etc.

Some Vocaloids have multipitch growl, if you try to use growl from a multipitch vocaloid, It will break your voicebank

Open your folder that contains the DDB Packer, and open a terminal.
you'll need to type the following command:

python mixins_ddb.py --src_path "path to your voicebank .ddi file" --mixins_path "path to the .ddi you want the growl from. --dst_path "the export path for the newly packed voicebank"


Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.
Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows
output

This creates a brand new voicebank. so you'll have to put these files where your VoiceDB is at, and **make sure they match the name of your VVD and other files in your CompID Folder.**

Replace growl to DDB with QuickBms

1. Pick out a V4 voicebank of your choice, as long as it has a growl it doesn't matter who it is.
2. Decompile your chosen one's .ddb using the quickbms script.

For this you'll need : Quickbms itself And this code (for decompiling) Select the code in quickbms and after doing so select the ddb you want to decompile

3. Find the growl wav files and replace them with your own. This is called a ddb injection, what you're doing is pretty much injecting your own samples into the ddb.

Make sure your growl wav file overlaps the voicebank's wav file correctly, just do it in Audacity or any audio editor software of your choice

4. After adding your growl, recompile the ddb.

For this you'll need to : Run the reimport2.bat file (it comes with quickbms) The use this And then you select the folder with your modified wav files.

5. Finally, using the ddb packer's mix-ins option, import the modified voicebank's growl into your homebrew.

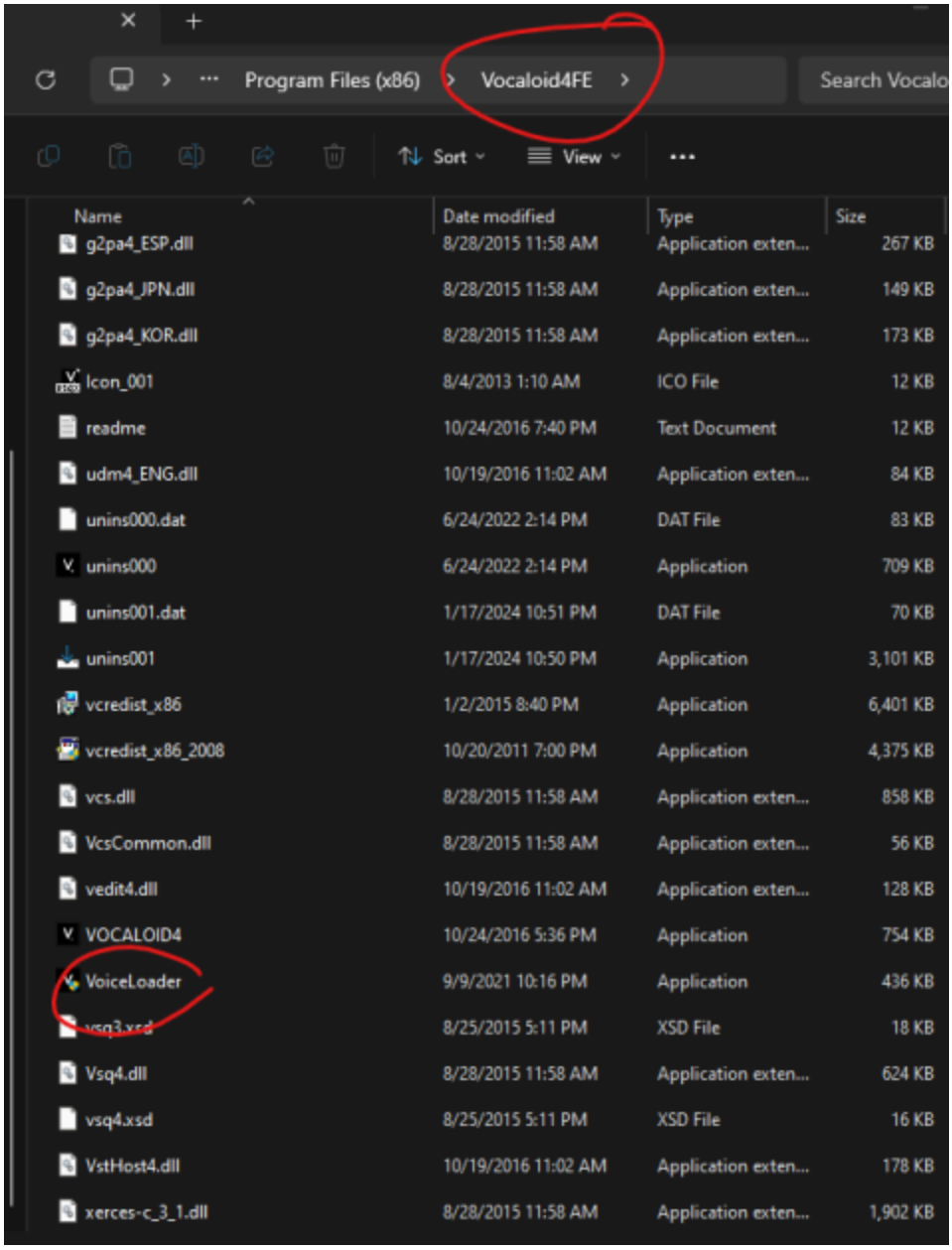
Here's a quick guide made by Daigasso!

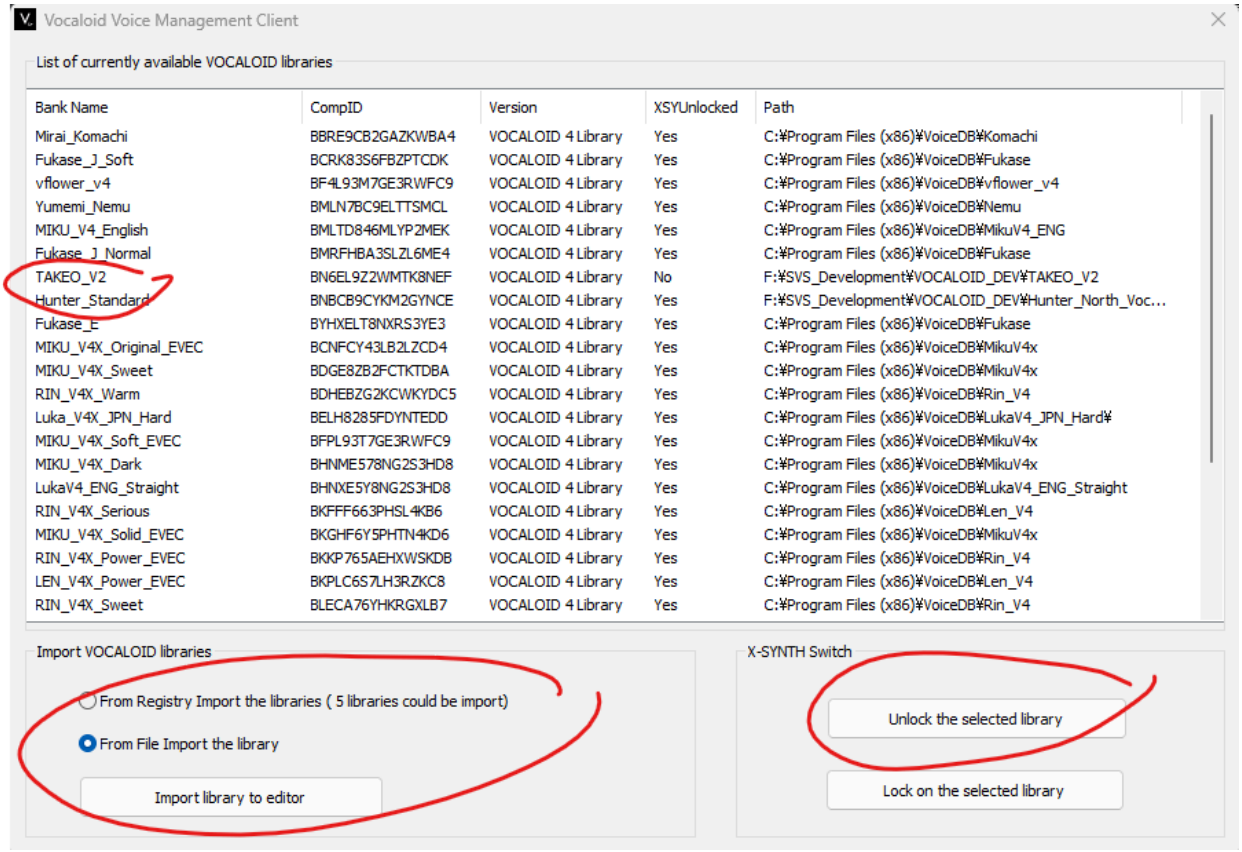
***If this is too hard, try the GUI. Be warned multipitch growl banks will break your bank.**
You are also able to use the GUI to add growl to V3 Voicebanks.

VOCALOID 4+ LOAD IN

DO NOT generate a whole new CompID, there's no point in that. Unless you're making a completely different bank. This Vocaloid already comes with a 'broken' XSY, so have fun crossing all the banks you want! Just remember the rules.

***For this you need VOCALOID 4FE+Alpha (Also works in Piapro V4x)**

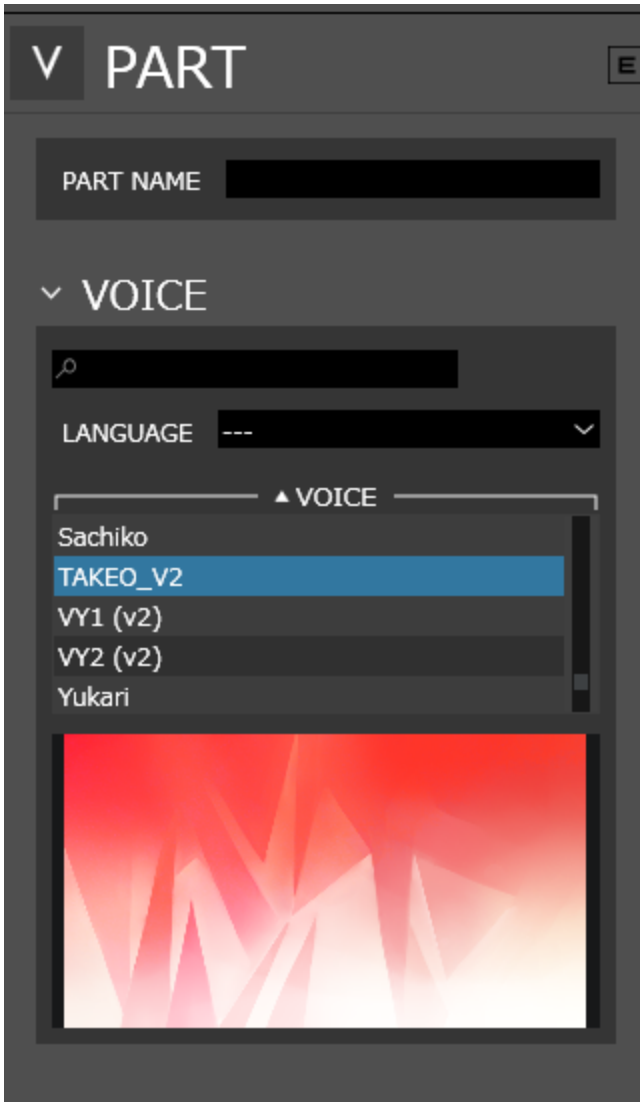

Program Files (x86)
Vocaloid4FE
Search Vocalo
Sort
View
Name
Date modified
Type
Size
g2pa4_ESP.dll 8/28/2015 11:58 AM Application exten... 267 KB
g2pa4_JPN.dll 8/28/2015 11:58 AM Application exten... 149 KB
g2pa4_KOR.dll 8/28/2015 11:58 AM Application exten... 173 KB
Icon_001 8/4/2013 1:10 AM ICO File 12 KB
readme 10/24/2016 7:40 PM Text Document 12 KB
udm4_ENG.dll 10/19/2016 11:02 AM Application exten... 84 KB
unins000.dat 6/24/2022 2:14 PM DAT File 83 KB
unins000 6/24/2022 2:14 PM Application 709 KB
unins001.dat 1/17/2024 10:51 PM DAT File 70 KB
unins001 1/17/2024 10:50 PM Application 3,101 KB
vcredist_x86 1/2/2015 8:40 PM Application 6,401 KB
vcredist_x86_2008 10/20/2011 7:00 PM Application 4,375 KB
vcs.dll 8/28/2015 11:58 AM Application exten... 858 KB
VcsCommon.dll 8/28/2015 11:58 AM Application exten... 56 KB
vedit4.dll 10/19/2016 11:02 AM Application exten... 128 KB
VOCALOID4 10/24/2016 5:36 PM Application 754 KB
VoiceLoader 9/9/2021 10:16 PM Application 436 KB
vsq3.xsd 8/25/2015 5:11 PM XSD File 18 KB
Vsq4.dll 8/28/2015 11:58 AM Application exten... 624 KB
vsq4.xsd 8/25/2015 5:11 PM XSD File 16 KB
VstHost4.dll 10/19/2016 11:02 AM Application exten... 178 KB
xerces-c_3_1.dll 8/28/2015 11:58 AM Application exten... 1,902 KB


Vocaloid Voice Management Client
List of currently available VOCALOID libraries
Bank Name | CompID | Version | XSYUnlocked | Path
Mirai_Komachi | BBRE9CB2GAZKWBA4 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥Komachi
Fukase_J_Soft | BCRK83S6FBZPTCDK | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥Fukase
vflower_v4 | BF4L93M7GE3RWFC9 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥vflower_v4
Yumemi_Nemu | BMLN7BC9ELTTSMCL | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥Nemu
MIKU_V4_English | BMLTD846MLYP2MEK | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥MikuV4_ENG
Fukase_J_Normal | BMRFHBA3SLZL6ME4 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥Fukase
TAKEO_V2 | BN6EL9Z2WMTK8NEF | VOCALOID 4 Library | No | F:¥SVS_Development¥VOCALOID_DEV¥TAKEO_V2
Hunter_Standard | BNBCB9CYKM2GYNCE | VOCALOID 4 Library | Yes | F:¥SVS_Development¥VOCALOID_DEV¥Hunter_North_Voc...
Fukase_E | BYHXELT8NXRS3YE3 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥Fukase
MIKU_V4X_Original_EVEC | BCNFCY43LB2LZCD4 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥MikuV4x
MIKU_V4X_Sweet | BDGE8ZB2FCTKTDBA | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥MikuV4x
RIN_V4X_Warm | BDHEBZG2KCWKYDC5 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥Rin_V4
Luka_V4X_JPN_Hard | BELH8285FDYNTEDD | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥LukaV4_JPN_Hard¥
MIKU_V4X_Soft_EVEC | BFPL93T7GE3RWFC9 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥MikuV4x
MIKU_V4X_Dark | BHNME578NG2S3HD8 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥MikuV4x
LukaV4_ENG_Straight | BHNXE5Y8NG2S3HD8 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥LukaV4_ENG_Straight
RIN_V4X_Serious | BKFFF663PHSL4KB6 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥Len_V4
MIKU_V4X_Solid_EVEC | BKGHF6Y5PHTN4KD6 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥MikuV4x
RIN_V4X_Power_EVEC | BKKP765AEHXWSKDB | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥Rin_V4
LEN_V4X_Power_EVEC | BKPLC6S7LH3RZKC8 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥Len_V4
RIN_V4X_Sweet | BLECA76YHKRGXLB7 | VOCALOID 4 Library | Yes | C:¥Program Files (x86)¥VoiceDB¥Rin_V4
Import VOCALOID libraries
From Registry Import the libraries (5 libraries could be import)
From File Import the library
Import library to editor
X-SYNTH Switch
Unlock the selected library
Lock on the selected library

***This will only have your voice show up in V4FE+A. If you wanna use it in 5-6 you're going to have to have an installer for your bank, install it and run Vocareg.**
You cannot cross 2 Voicebanks of a different language.
You cannot cross 2 Voicebanks of a different format.
You cannot cross CVVC only banks. (To circumvent this just do one VCV segment ([V C V]) during development.)

Add a portrait to Vocaloid ⅝

Go to
C:\Program Files\Common Files\VOCALOID5\Resource\Voice
C:\Program Files\Common Files\VOCALOID6\Resource\Voice
Here you will see multiple CompID's. **Copy your ID and make a new folder, then paste it onto the folder.** Next you can throw any random photo in there. Use someone else's as reference for the dimensions if you need to (600x324). Make it a BMP and change the name to Setup. Then your photo will show up.



FAQ

Q: How hard would you say this is?

A: This is like an amped up UTAU Creation. Vocaloid is not forgiving at all about marker placement. One pitch can either help or completely destroy your bank. If you're not careful about how and where you pull your samples you can end up with a very damaged bank. No one wants that for you. Overall you just gotta fuck around and find out.

Q: Can I make V6AI?

A: **NO! Don't be fucking stupid.**

Q: CVVC vs VCV?

A: Real talk? You'll have an easier time with CVVC banks, and produce about the same sound. So, just add that 1 VCV plug and go on. Do VCV for needing things like, some inbetweens or openings. You cannot do vcv alone, you need cvvc plugs to stack!

Q: Is blank how i record blank?
A: The transcripts hold all the answers. Please look at them when you are confused..

Q: I have the pitches, but I don't know how they'll work?
A: Deepvocal! DV is actually pretty accurate when it comes to this. So doing a quick test with DV wouldn't be bad for seeing a timbre. Just use your 'a' samples.

Q: Should I clean my samples before labeling?
A: I personally say yes, but As long as you clean the samples before you add it to your bank. (Meaning you can completely label them and come back, because the DevKit is real time.) You'll be fine.

Q: Are all these samples necessary?
A: Yes, yes they are. There's no way around it. You either do all of them or- you work phonetically. Vocaloid is very specifically built! One off thing, it can mess you up real bad. You think you've seen finicky just wait!

Q: What do I do if the DB tool crashes?
A: The devkit saves in real time, just reopen it.

Q: Okay so I have this UTAU, and I wanna port it. possible?
A: Very possible! Will your Vocaloid come out slightly weird? Yes because you don't have every sample. YOU CAN however cheese it by using CVVC strings for all the N's, and forcing the [Sil C] and [C Sil].Then ya copy and paste a few things, and we have a bank! (Don't worry Miku's banks re-use samples all the time! In Fact Vy1V4 natural is the 3 other banks in one bank!)

Q: Should I work on this on an External drive?
A: Vocaloids can either come out to be 1 GigaByte or 20! And that's in dev form. Work on an external drive.

Q: What do I do when I don't have a premade dictionary?
A: Download one from below, build it yourself, or just straight up cheese it. Now Vocaloid only supports the languages it supports, but that doesn't stop you from adding in your phonetics however you want. Do not add C/V_0 for V4, because that will already work. All you have to do is type it in phonetically.

Q: Why doesn’t [-] work?
A: Not sure why this happens? You might have to remake your bank. As in, new bank and then re-add all the samples.

Q: Can banks work in Piapro v4x
A: Yes. You can load them the exact same way you do in V4FE+A. You can even add custom icons!

Q: I’m getting a weird noise when I use a certain phonetic, what's up?
A: Your markers are placed wrong! Vocaloid is not forgiving in the slightest. The devkit is held together by a piece of tape sure, but you have to place your markers correctly. Or they have a pitch error. In the main window use control+click to correct the pitch of the sample. control+D+click to delete pitch.

Q: Can I use my bank in tunelab?
A: Install it properly and the sky's the limit. (Meaning when you have your installer run it as an admin and it’ll register no key needed.)

Q: Can I share my bank?
A: Yeah, give it out with the VVD, and compID and others can use it no problem. Give it out with the installer and they can use it in tunelab! Treat’em like UTAU. Make sure you use discretion, and don’t call it official!

Q: Can I Jinkiri, or RVC?
A: Again, treat it like an UTAU. Tho I don't encourage either.

Q: The packer gave me -9 or -1? (Errors in packer)
A: -9 is a segmentation error; usually occurs when your configuration is either incomplete or not done correctly/properly. -1 i BELIEVE is an audio format error (i'm not exactly sure though i think this is it) - *daigasso*

Q: The hardest part is?
A: When you have no transcripts and have to do it all by hand. When you’re a standalone dev! So you have to do all the labeling and troubleshooting alone. It’s better to work in groups for things like these. No part of this is easy, and there’s a huge learning curve if you don’t understand vocal synth at all! Honestly learn how to properly configure an UTAU, CVVC and VCV. Then come back! And do the Vocaloid.

Q: What’s up with the N\?

A: Vocaloid relies on context <3! N\ is just one of the many N phonetics. So you need to make sure you get to all of them okay <3 Even the k' in kya is contextual.

Q: Does the '?' do anything?
A: You can use this to get a natural sounding stop. For example 'Steppu', The ? is the 'p' between e and pu or just any aspiration. In truth, ? can be vocal fry and you can use 'asp' for aspirations!

Q: It says BAD in the config?
A: Your markers are overlapping, or a phonetic isn't in the dictionary. Sometimes it doesn't matter.

Q: Why is my audio glitched out?
A: Check your format. Then re-export.

Q: What's up with Vocaloids scales?
A: They're off by 1, C4 is C5.

Q: Dev Editor doesn't detect my bank?
A: Make sure the paths are both correct. The only difference is one should have .tree at the end

Q: How do I EVEC?
A: Add more vowels using the dictionary, go from there.

DISCLAIMER

We're not liable for what you do at your own discretion. Do not try to profit off the DBTool or Vocaloid.
Do not steal voices and make banks with them.

RESOURCES

VocaloidDBTool3 Here
Vocaloid4Dev Here
SMSTools2 Here

Vocaloid 4FE+Alpha Here (This includes a tool to register packed voice banks without using the installer. However the voice will only show up in V4+A editor)
Piapro V4x Here (This includes a tool to register packed voice banks without using the installer. However the voice will only show up in Piapro)
Deepvocal Here Well, test your range buddy!!
Canned Bread's Swiss Army Knife Here (This tool allows you to create Blank .trans files, and converts .oto to .seq files.)
QuickBMS Here
VVD Editor Plus Here (Allows you to generate VVD's, Install & Uninstall scripts for your voicebank. Useful when you need to pack the VB for use on other editors.)
DDB Tools Here (Used for packing the Voicebank in a readable format that V3, V4, V5 & V6 can use. You must supply your own ddb & ddi for growl. we cannot provide that to you* Use VY1's or Flower's. CyberSongman is the original growl.)
Colab Version Here (Updated to use Github repository)
DDB Tools GUI *(From UselessBug)* Here An easier way to pack your voicebank. (You need python for this. To open it click your address bar and type cmd, Then do python -m pip install -r requirements.txt and then
python GUI.py. Once installed you'd only need to do python GUI.py! YOU NEED TO DO CMD EVERYTIME CMD>python GUI.py! Everytime!)
Oto2seg Here Converts your UTAU oto to Vocaloid ready files. It will split your samples into small pieces. Make sure to have ffmpeg installed, and download pydub then throw it in the same folder. It's not perfect so be sure to remain clean. For sationaries, and C/V Sil samples you're your own. m m' N N' and J samples also don't autogen. IT WILL skip over uncommon japanese configures/Labels like Hu(ほぅ), Fu(ふぅ), Wu(うぅ), and Yi(いぃ).
Custom WordBank Base Here (This won't completely work like another word bank. To use it you have to replace one of the dictionaries in the Vocaloid folder, making sure that both files are the same name as the one you're replacing. Be sure not to delete the original word bank for normal use.)
Gumi English List Chart (not base) Here If you can convert this yourself you'll have a base
Japanese CVVC/VCV Reclist+TransFile *(From Enlilgen)* Here Not perfect at all
Japanese CVVC Reclist+Transfile *(From Inochi-P)* Here
Japanese word based list *(From Ko Ko)* Here Record it like an utau. 1 syllable is one clap. (N is it's own syllable) Not perfect but a good base. You will need to copy and paste your samples
English word based list *(From Canned Bread/luciozundiezz)* Here Record it like an utau. 1 syllable is one clap. (Reference gumi's chart for which samples you need to make copies of) Not perfect but a good base. You will need to copy and paste your samples

Spanish word based list pdf *(From AnotherNN)* Here Not perfect, but trimmed.
Vocaloid2Keygen Here
Dictionaries *(From daigasso)* Here
Melodyne Here For editing or correcting pitch of samples, if needed!
Takeo V2 (Example CVVC bank+Dev files) *(From Enlilgen)* Here (Updated with louder and cleaner samples) I give this for you to learn and answer your questions on what to do and not to do. Take this and dissect it to your heart's content, just follow the readme! Experiment and all, go crazy, use it as a basis for developing your own reclist or banks.
Takeo V4 HARD (Example VCV/CVVC bank+Dev files) Here Here The same rules apply here from the V2 bank. (You can cross both these banks)

For questions or help go to @Enlilgen on Twitter.

#jinkoloid, so we can see it

Join the server on discord

TRANSLATIONS

Translated versions of this document

translations are from other users, may be out of date or altered

Chinese Translation Here - By Slidingwall
Portuguese Translation Here - By Bread